Journal of Computers Vol. 33 No. 5, October 2022, pp. 49-60
doi: 10.53106/199115992022103305005

A Fast Clustering Method of KPI Data Based on IP-Kshape Algorithm

Yun Wu[1], Yu Shi[1], Jie-Ming Yang[1], Zhen-Hong Liu[1*], Li-Shan Bao[2], Chun-Zhe Li[3]

[1] School of Computer Science, Northeast Electric Power University, Jilin 132012, China
{838558160, 1430802937, 670172713, 33249648}@qq.com
[2] Management Consulting Branch of Jiangsu Xinshun Energy Industry Group Co. Ltd, Nanjing 221000, China
138013800@qq.com
[3] Tonghua Power Supply Company, State Grid Jilin Electric Power Co. Ltd, Tonghua 134000, China
3337247@qq.com

Received 15 November 2021; Revised 2 February 2022; Accepted 3 March 2022

Abstract. Aiming at the problem of large amount of intelligent operation and maintenance KPI data and poor clustering effect, this paper proposes a fast KPI clustering method based on the IP-Kshape (IAE-PAA-Kshape) algorithm. First design an improved autoencoder algorithm (IAE), add a convolutional layer and a two-way LSTM network layer to the standard autoencoder, to achieve smooth denoising of KPI data and timing feature extraction; then the KPI data features are clustered based on the PAA-Kshape algorithm, and the PAA algorithm is used to perform dimension compression and Kshape algorithm to solve the drift problem of KPI sequences, which improves the clustering speed and accuracy of KPI data. Through experimental comparative analysis, it is proved that the method proposed in this paper can better realize the rapid clustering of KPI data, and the time efficiency and accuracy are better than traditional machine learning or deep learning methods.

Keywords: AIOps, KPI clustering, Kshape algorithm

1 Introduction

In the era of big data, more and more companies place their main business or transaction data on servers for operation and maintenance. With the development of Internet software and hardware, the current automated operation and maintenance has gradually shifted to intelligent operation and maintenance (Artificial Intelligence for IT Operations, AIOps). The concept of intelligent operation and maintenance was proposed by Gartner in 2016, aiming to use big data, machine learning and other methods to improve operation and maintenance capabilities [1]. In enterprise intelligent operation and maintenance, in order to provide reliable and efficient services to tens of millions or even hundreds of millions of users, operation and maintenance personnel usually use Key Performance Indicators (KPIs) to monitor the service performance of various applications. These KPIs are roughly divided into two types: service KPIs and machine KPIs. Service KPIs refer to performance indicators that can reflect the scale and quality of web services, such as web response time, web page visits, number of connection errors, etc. Machine KPIs refer to performance indicators that can reflect the health status of machines (servers, routers, switches), such as CPU usage, memory usage, and network card throughput [2].

Most of the operation and maintenance KPIs are expressed in the form of time series and are closely related to enterprise business. They have the characteristics of periodicity, multi-dimensionality and variability. Its various application analysis modules (such as anomaly detection [3-4], bottleneck analysis) and root cause location [5-8], etc.) It is also necessary to select an appropriate clustering algorithm according to the shape and trend of different types of monitoring indicators, that is, a clustering method with good effect is required to classify KPIs. There have been some clustering methods with better performance, such as YADING [9] KPI clustering through PAA dimensionality reduction and DNSCAN density clustering, ROCKA [10] KPI clustering through baseline extraction and DBSCAN density clustering, etc. However, there are still many difficult problems in KPI clustering. For example, the time span of KPI data is too large, often including tens of thousands of data points, and the fluctuations are large, and the trend of change is difficult to find; KPI data will change shape over time; due to system or man-made reasons, the trend will have a certain degree of left-to-right deviation, which will cause a certain misjudgment in the cluster analysis of KPI.

In order to solve the above problems, this paper proposes a fast clustering method of KPI data based on IP-Kshape algorithm. First, design an improved autoencoder (IAE) structure for feature extraction of KPI data, and

* Corresponding Author

mainly add three specific hidden layers to the standard autoencoder to achieve smooth denoising and timing feature extraction of KPI data; then the PAA-Kshape clustering model is proposed, which uses the PAA algorithm to compress the sequence length of KPI data, and then uses the SBD distance in Kshape to solve the problem of large clustering errors caused by drift between KPI data, and realizes the KPI data fast and accurate clustering.

2 Related Work

2.1 Autoencoder

Autoencoder (AE) [11] is an unsupervised learning technique that uses neural networks for representation learning. In other words, we design a neural network architecture that imposes a "bottleneck" in the network and forces the original input to compress the knowledge representation. If the input features are independent of each other, this compression and subsequent reconstruction will be a very difficult task. However, if there is a certain structure in the data (that is, there is a correlation between the input features), this structure can be learned and used when forcing the input to pass through the bottleneck of the network.

The Denoising Autoencoder (DAE) is based on the autoencoder and learns some useful information by changing the reconstruction error term of the loss function. Mix noise data in the training set, and make the autoencoder learn to remove this noise to obtain a real input that has not been polluted by noise. Therefore, this forces the encoder to learn to extract the most important features and learn more robust representations in the input data, which is why its generalization ability is stronger than that of general encoders.

2.2 PAA

Aggregate query is a commonly used but time-consuming database operation. Compared with accurate query, it is usually a better choice to return approximate results that meet the confidence interval to the user with much less response time. The existing approximate query methods cannot efficiently process approximate aggregate queries that meet arbitrary precision on massive data. Some scholars have proposed a new algorithm PAA (partition-based approximate aggregation) [12-13] to effectively deal with the approximate aggregation that satisfies any confidence interval. The data space of dimensional attributes is divided into spatial regions of the same size, and each segment maintains tuples of dimensional attributes that fall into the corresponding spatial region.

The PAA algorithm maintains a random sample RS of the table, and its execution includes two stages. In stage 1, if the pre-built random sample RS cannot return an approximate result that meets the user's requirements, then in stage 2, the PAA algorithm obtains more random tuples from the shard set IPS corresponding to the spatial region intersecting with the query area. The characteristics of the PAA algorithm are: how to obtain the required random tuples from the IPS without knowing the number of tuples that each segment contains in the IPS satisfies the predicate; how to effectively reduce the random I/O cost in stage 2. Experiments show that compared with the existing methods, the PAA algorithm can achieve a speedup of two orders of magnitude.

2.3 Kshape

Most time series analysis methods [14-15], including clustering algorithms, rely on the choice of distance measurement. When comparing two series, the key question is how to deal with the distortion problem, which is also a characteristic of time series. Ideally, a shape-based clustering algorithm divides time series into the same cluster based on shape similarity, rather than the difference in magnitude and stage. The existing shape-based methods have two main drawbacks: (a) These methods cannot be extended to large data sets, because these methods are time-consuming for calculation or distance measurement. (b) The effectiveness of existing methods is limited to specific fields or data sets. Moreover, these algorithms are not compared with classic ones such as partition clustering.

The Kshape method proposed in [16] is similar to k-means but has obvious differences. The method of calculating cluster centers and the distance measurement of the Kshape method are different from those of k-means. Kshape tries to preserve the shape of the time series when comparing. Therefore, Kshape needs a distance measurement method for the invariance of scaling transformation. Unlike other clustering algorithms, Kshape uses a standard cross-correlation distance measurement method. This algorithm can efficiently compare sequences and calculate sequence centers while ensuring zoom invariance, translation invariance and transformation invariance, so it is more suitable for time series clustering processing.

3 Primary Coverage

3.1 Feature Extraction Model Based on the IAE Algorithm

Since KPI data comes from real business scenarios, and with the passage of time and the gradual increase in the number of users, the amount of data generated is increasing exponentially, which may lead to various unexpected situations, for example, there will be gaps and exceptions in the data. In order to reduce the work pressure of enterprise operation and maintenance staff, this article aims to design an intelligent KPI data feature extraction model, which can accurately extract the timing features of KPI sequence data while solving the above-mentioned KPI vacancies and abnormalities.

First, preprocess the KPI data. For data missing due to unpredictable risks in the business scenario, linear interpolation functions are used to complete the missing values; for outliers, this article sets data points that are more than 95% away from the average of the KPI data as outliers, to replace the original data points with linear interpolation.

Then perform feature extraction on KPI data. This paper designs an improved autoencoder structure (IAE), that is, three specific hidden layers are added to the standard autoencoder. In the encoding process, the first layer is set to 1D convolution, the second layer is the forward LSTM to extract the timing features, and the third layer is the reverse LSTM to extract the corresponding timing features. In the decoding process, the network structure is symmetrical to the coding layer, and the root mean square error is used as the loss function during network training. Put the preprocessed KPI data into the model for network training, complete the preliminary dimension reduction and denoising of the KPI data, and realize the feature extraction of the KPI data.

3.2 The KPI Fast Clustering Model Based on the PAA-Kshape Algorithm

Due to the wide range of KPI data sources, the data increment is quite large, and the data will not only produce vacancies and abnormalities, but also other problems such as drift and overlap. In response to the above problems, the PAA-Kshape clustering algorithm proposed in this paper first performs segmentation and aggregation of KPI data, that is, the data points in a small area are replaced with their average values to reduce the length of time series data and minimize the overlap of KPI data; then perform Kshape clustering on the aggregated data, classify KPIs with the same shape and the same timing characteristics into one category, and solve the drift problem of KPI data based on the characteristics of the SBD distance.

PAA Segmented Aggregation. After KPI feature extraction after IAE, the data still has the problem of complex sequence length. This paper uses PAA algorithm to simplify KPI data. The segmented aggregation of time series data through the PAA algorithm can significantly reduce the volatility complexity of the KPI curve.

Segmented Aggregate Approximation (PAA) is a feature extraction algorithm based on the average value of segmented sequences. PAA first divides the time series into equal-length segments, then calculates the average of these segments, and uses the average to represent the value of all data in the segment. Suppose the length of the time series $X = \{x_1, x_2, x_3, ..., x_n\}$ is n, and the length is reduced to d dimension after using the PAA algorithm, which can be represented by $\overline{X} = \{\overline{x}_1, \overline{x}_2, \overline{x}_3, ...\overline{x}_d\}$, where [13]:

$$\overline{x}_i = \frac{d}{n} \sum_{j=\frac{d}{n}(i-1)+1}^{\frac{n}{d}i} x_j \quad . \tag{1}$$

After performing PAA dimensionality reduction, subsequent calculations are performed in a lower dimensional space, which reduces the time complexity; in addition, for each segment, the average value is substituted to smooth the noise to varying degrees. However, PAA also has its drawbacks. Since each segment is represented by its average value, it will lose some of the more critical information in the time series, such as maximum values, minimum values, and other important forms. In addition, the information of the sequence after PAA dimensionality reduction depends on the reduced dimensionality d. The smaller the value of d, the rougher the time series represented by PAA, the more information is lost, and the greater the degree of dimensionality reduction; on the contrary, the finer the time series represented by PAA, the less information is lost, and the smaller the degree of dimensionality reduction. Therefore, when using PAA for dimensionality reduction, the trade-off between the quality of the sample after dimensionality reduction and the number of segments d should be considered. The value of d needs to be determined according to $d = \frac{n}{window\_size}$.

Kshape Fast Clustering. After feature extraction and dimensional compression of KPI data, this section formally clusters KPI data. Since KPI data is the monitoring data of enterprise server operation and maintenance, it has the characteristics of time series and has a certain degree of periodicity, so this article will use the Kshape clustering method based on shape similarity to cluster KPI data. Kshape uses SBD distance as a similarity measurement method, see formula (2), which can solve the possible left-right deviation and low time efficiency of KPI data, and complete the cluster analysis of KPI time series data [16].

$$SBD(x, y) = 1 - \max_w \left(\frac{CC_w(x, y)}{\sqrt{R_0(x, x) \cdot R_0(y, y)}}\right) . \tag{2}$$

The value range of the SBD distance is [0, 2], and 0 indicates that the two time series are the most similar, and the efficient calculation of the SBD can be seen from this equation:

$$CC_w(x, y) = R_{w-m}(x, y), w \in 1, 2, ..., 2m-1 . \tag{3}$$

It can be seen that the time complexity of $CC_w(x, y)$ is $O(m^2)$, where m is the length of the time series. Because the calculation process of cross-correlation is very similar to the calculation process of convolution, according to the convolution theorem, the convolution of two time series can be calculated as the Inverse Discrete Fourier Transform (IDFT) of the product of the Discrete Fourier Transform (DFT) of a single time series, where DFT and IDFT are shown in formula (4) and formula (5) respectively [16]:

$$F(x_k) = \sum_{r=0}^{|x|-1} x_r e^{\frac{-2jrk\pi}{|x|}}, k = 0, 1, ..., |x|-1 . \tag{4}$$

$$F^{-1}(x_r) = \frac{1}{|x|} \sum_{r=0}^{|x|-1} F(x_k) e^{\frac{jrk\pi}{|x|}}, r = 0, 1, ..., |x|-1 . \tag{5}$$

Where $j = \sqrt{-1}$. If a sequence is first inverted by 180 degrees in time, the calculation of cross-correlation is the convolution of two time series, which is equivalent to taking the complex conjugate in the frequency domain (denoted by *), as shown in formula (6) [16]:

$$CC_{w(x,y)} = R_{w-m}(x, y), w \in 1, 2, ..., 2m-1 . \tag{6}$$

$$CC(x, y) = F^{-1}\{F(x) * F(y)\} . \tag{7}$$

However, the calculation of DFT and IDFT still requires $O(m^2)$ time. By using the fast Fourier transform algorithm, the time complexity can become $O(m \log m)$. The recursive algorithm calculates the FFT by dividing the FFT into power-of-two blocks. Therefore, in order to further improve the calculation performance of the FFT, when $CC\ (x, y)$ is not an exact power of 2, you can fill in 0 between x and y so that it is still the length of the power of 2.

The Kshape algorithm can efficiently compare sequences and calculate sequence centers while ensuring zoom invariance, translation invariance and transformation invariance. In each iteration, the K-shape algorithm performs two steps:

(1) In the allocation step, the algorithm compares each time series with all calculated geological centers, and assigns each time series to the cluster closest to the centroid to update the membership in the cluster;

(2) In the refinement step, the cluster centers are updated to reflect the changes in the cluster members in the previous step.

The algorithm repeats these two steps until the cluster membership does not change or the maximum number of iterations allowed is reached. It can complete the rapid clustering of KPI time series data.

4 Experimental Analysis

In order to verify the effect of the IP-Kshape based KPI data fast clustering algorithm proposed in this article, this article uses the operation and maintenance data set from Alibaba to conduct related experimental analysis. The data size is 4023*10000, and the row attributes of the data indicate different KPI samples, column attributes are time points, that is, the KPI data in this experiment is the continuous time series values of different KPI samples over a period of time.

4.1 Experimental Environment

Hardware environment: Cuda9.0-GPU-32.00GB (RAM)-700.00GB (SSD);
Software environment: Linux-Ubuntu 18.04-PyCharm Community 2021- Python 3.8.

4.2 Feature Extraction

In the experiment, the original KPI data was first processed with missing values, the missing values or data points more than 95% from the average value were supplemented and replaced by linear interpolation, and then the feature extraction based on the improved autoencoder was performed to analyze the KPI data. A certain degree of smoothing is carried out to facilitate the related experimental analysis in the later stage.

4.3 IAE-PAA-Kshape Cluster Analysis

The PAA-Kshape clustering algorithm proposed in this paper first performs segmental aggregation processing on KPI data, that is, the data points in a small area are replaced with their average values to reduce the length of time series data points; then the aggregated data is Kshape clustered. Group KPIs with the same shape and timing characteristics into one category. In order to compare the effects of different clustering methods, this section will use the Kshape method, PAA-Kshape method, and IAE-PAA-Kshape method to analyze the calculation examples:

Kshape Algorithm. Use the Kshape method to cluster the KPI data set. After many experiments and analysis, it is found that the best effect is when the number of clusters is 4, so this parameter is set as the running parameter of the model. The running result is shown in Fig. 1, and the cluster centroid is shown in Fig. 2.

It can be seen from Fig. 1 and Fig. 2 that the values and fluctuation trends of the first three time series graphs are relatively similar. The fluctuation ranges of cluster1 and cluster2 are more similar, while the fluctuation range of cluster3 is slightly larger than that of the first two and relatively intense. The time series curve data of Cluster4 is close to 0, and the fluctuation is more regular. It can be clearly seen that this type of curve has a certain periodicity. In short, it can be clearly seen that the clustering effect of this method is not very good, especially the first three categories seem likely to be the same category.

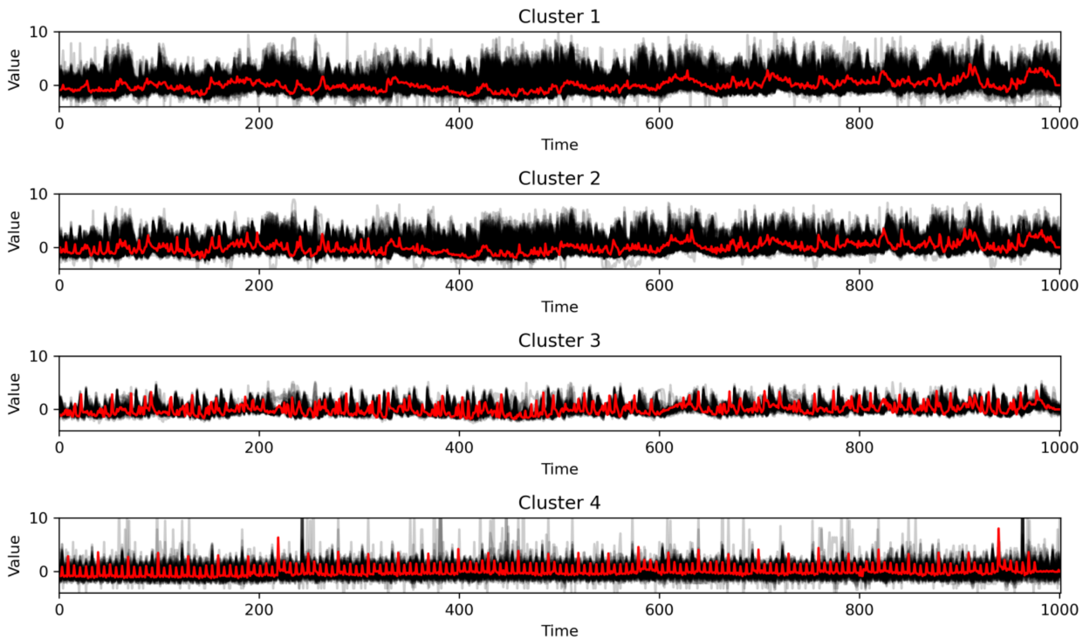

Fig. 1. The results of the Kshape algorithm clustering

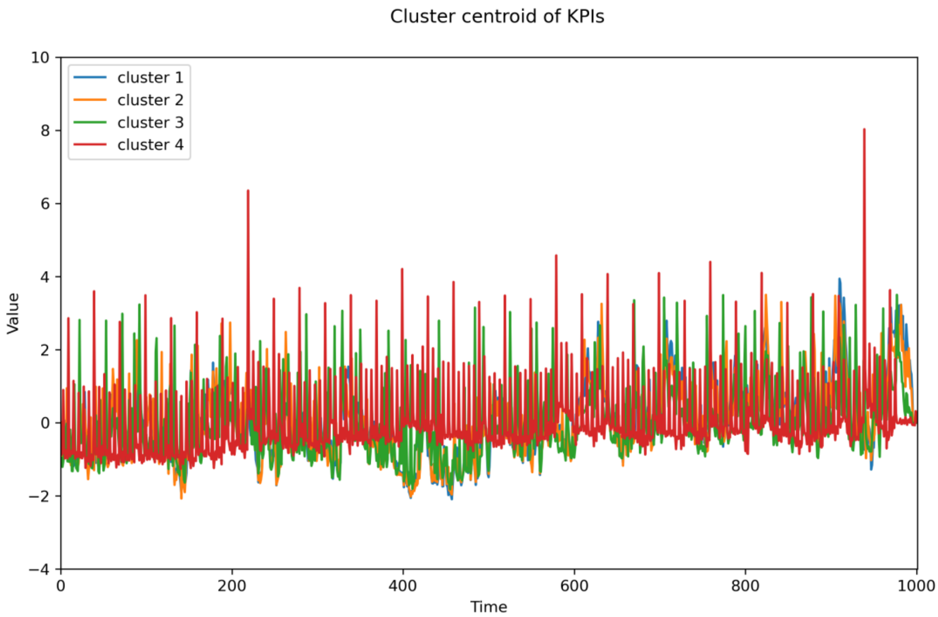

Fig. 2. Display of the centroid clustering results of the Kshape algorithm

PAA-Kshape Algorithm. The visualization results given by the above Kshape method are complex, and it does not allow people to see obvious differences between different categories at a glance. Therefore, this section uses the PAA-Kshape clustering model to cluster the KPI data set. First, the KPI data is aggregated through the PAA algorithm. Among them, according to multiple experiments, it is found that the clustering effect is best when window_size=20. Therefore, in this experiment, the window_size parameter in the PAA-Kshape algorithm is set to 20, and the result is shown in Fig. 3. It can be seen that PAA has shortened the upper and lower ranges of the original data and retains the fluctuation trend of the KPI curve.

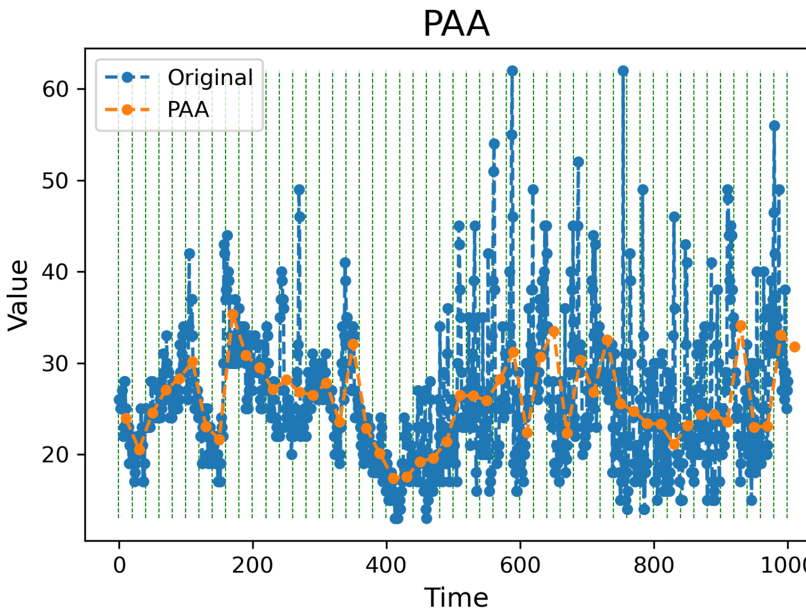

Fig. 3. The KPI results after PAA treatment

After the KPI data were processed by the PAA algorithm, the KPI was clustered using the Kshape algorithm, the parameters remained consistent with the previous experiment, and the resulting clustering results and centroid results are shown in Fig. 4 and Fig. 5.

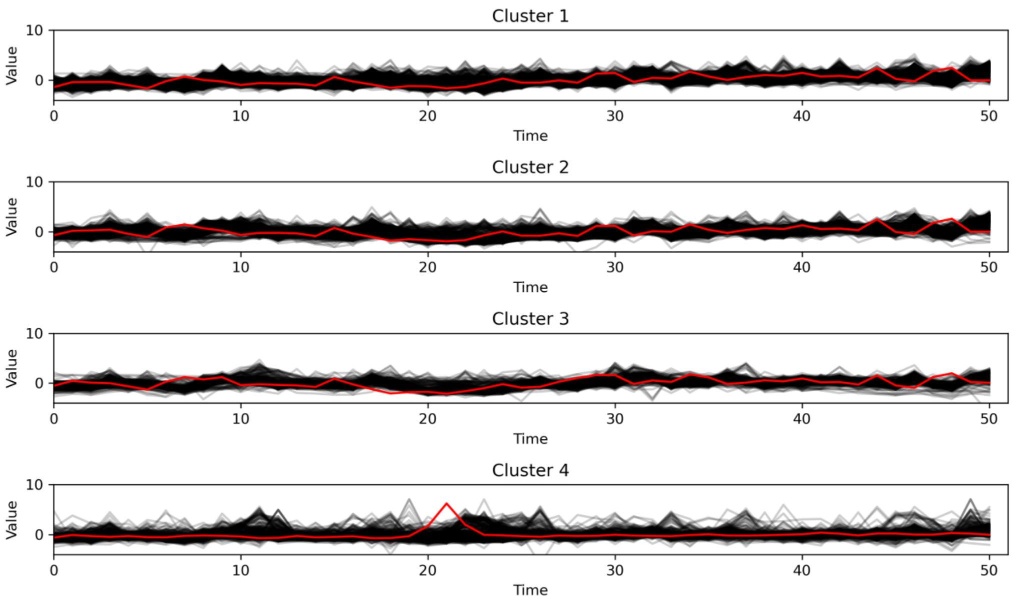

Fig. 4. The clustering results of the PAA-Kshape

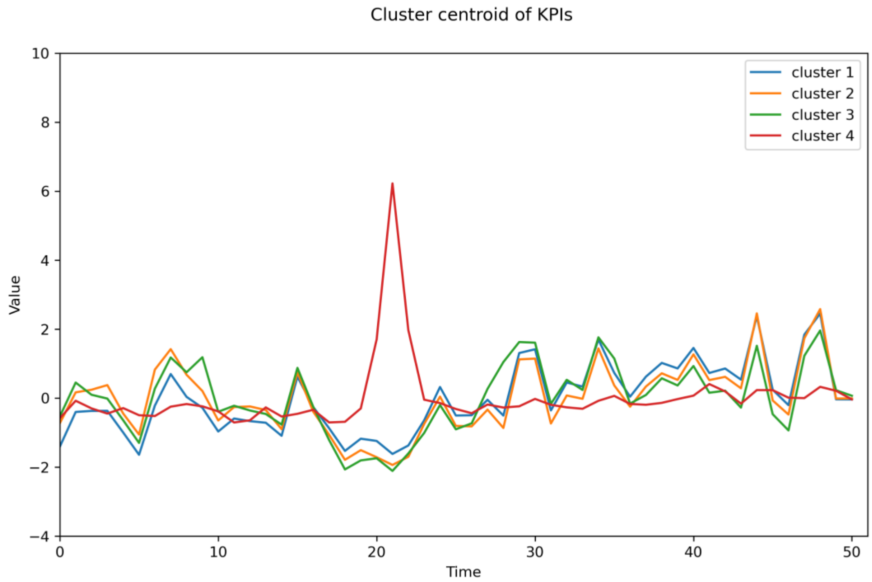

Fig. 5. The PAA-Kshape cluster centroid

As seen from Fig. 4 and Fig. 5, the timing curves of the first three categories are almost consistent after the PAA-Kshape clustering algorithm, which then yields a possible error in accuracy for KPI clustering only through the standard Kshape algorithm. Clustering results, after being adjusted for the clustering parameters, and the centroid of the PAA-Kshape algorithm are shown in Fig. 6 and Fig. 7.

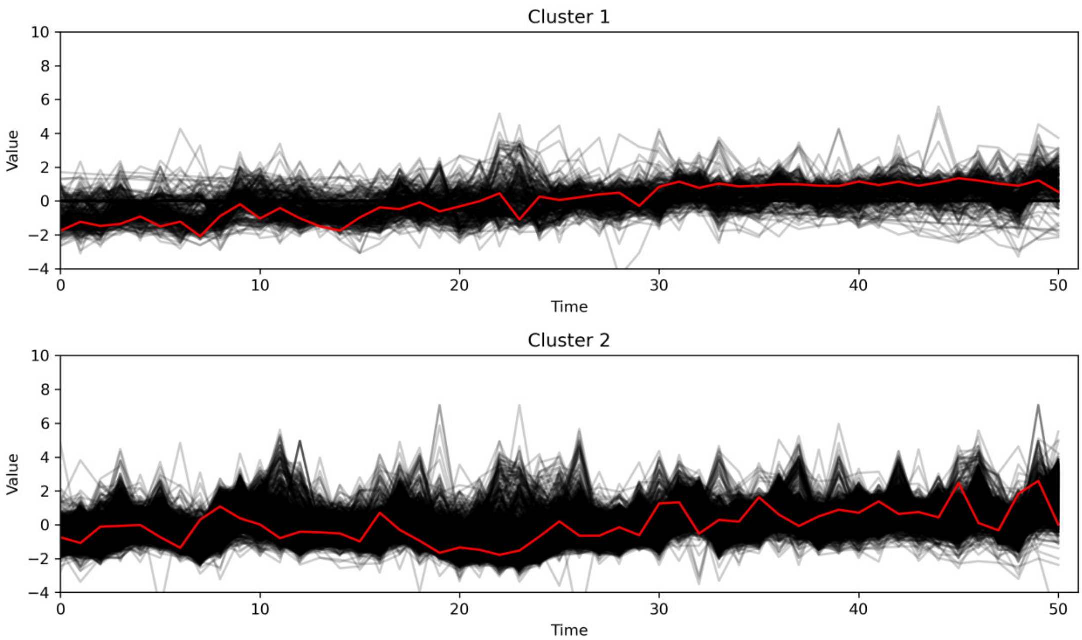

Fig. 6. The clustering results of PAA-Kshape when the number of clusters is 2

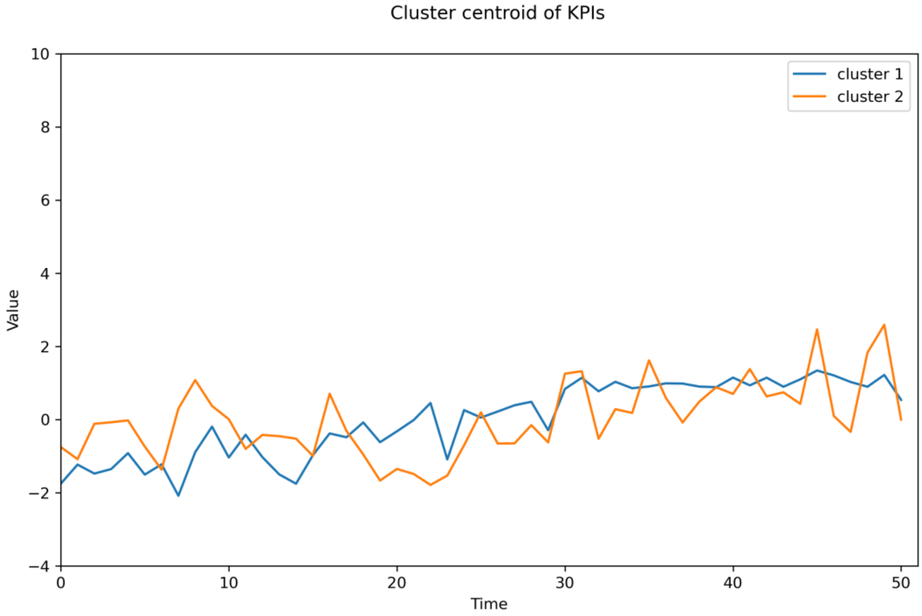

Fig. 7. PAA-Kshape cluster centroid when the number of clusters is 2

It can be seen from Fig. 6 and Fig. 7 that the effect of the PAA-kshape clustering algorithm has achieved relatively good results. Compared with the previous experiment where the number of clusters is 4, the result of this experiment is that the number of clusters is 2, and the visualization effects of the two results have been significantly different.

IAE-PAA-Kshape Algorithm. It can also be seen from Fig. 6 that there is still a big difference between the cluster centroid and other similar KPI time series curve changes and fluctuations. This paper uses the IAE algorithm on the original KPI data to simplify and smooth the original KPI data to a certain extent, then extract the time series characteristics of the KPI data, and then train and test the clustering model of PAA-Kshape to obtain the test results. As shown in Fig. 8 and Fig. 9.

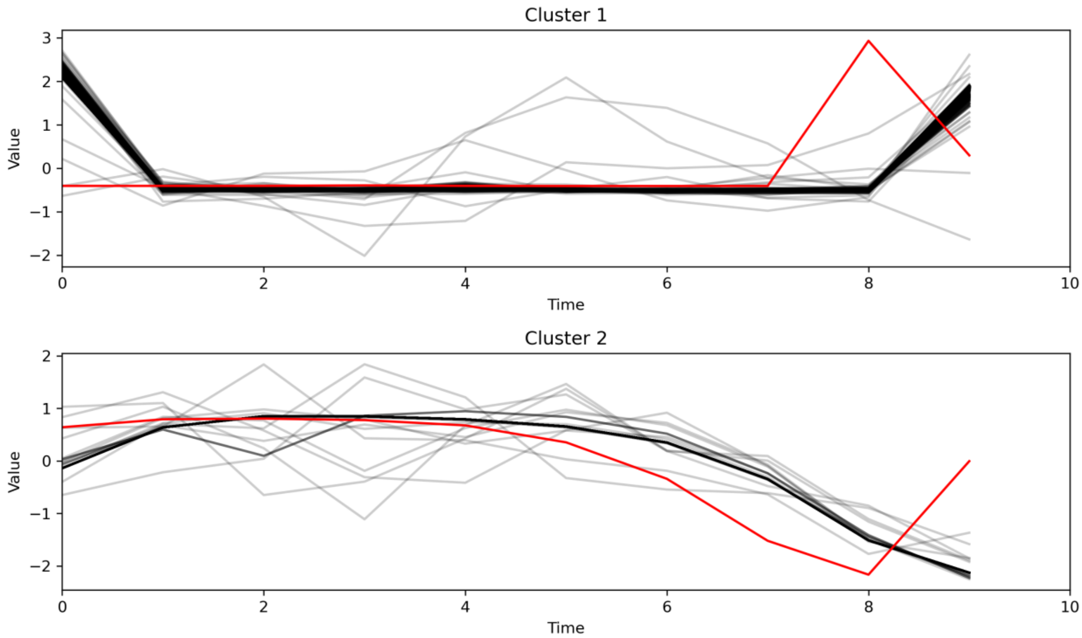

Fig. 8. The clustering results of IAE-PAA-Kshape

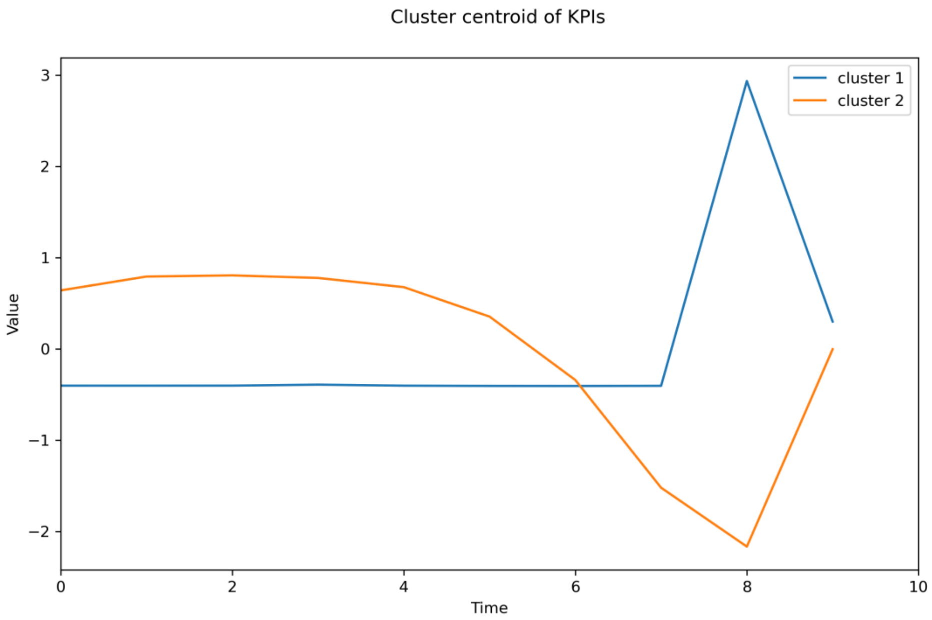

Fig. 9. IAE-PAA-Kshape cluster centroid

Obviously, it can be seen from Fig. 8 that the gap between the KPI data of different categories and their centroids has been reduced a lot, which can almost be regarded as overlapping, and the length of the time series has also been simplified from 50 to 10 units, indicating that IAE has achieved good results in the preprocessing and feature extraction of the original KPI data, and also obtained a good visualization effect. It can be seen from Fig. 9 that the two categories of KPI have obvious fluctuation characteristics, which can be displayed to the enterprise operation and maintenance personnel in a more simplified way, so that they can conduct in-depth mining of related data later.

In summary, IAE-PAA-Kshape (IP-Kshape) is better than the pure Kshape and PAA-Kshape methods in terms of clustering accuracy and speed, and has a good visualization effect.

4.4 Analysis of Cluster Evaluation Index

In order to better demonstrate the effectiveness of the method proposed in this article, this section compares and analyzes the three methods of Kshape, PAA-Kshape and IAE-PAA-Kshape. The specific effects are shown in Table 1.

In this paper, the distance between each sample and its nearest cluster center is used as the evaluation index of KPI cluster analysis, as shown in formula (8). Among them, m is the number of samples, n is the number of cluster centroids, and dist represents the Euclidean distance. The smaller the value of Error, the smaller the clustering error and the better the clustering effect.

$$Error=\sum_{i=0,j=0}^{i=m,j=n} \mathrm{dist}(\mathrm{x}_i, c_j) \ . \quad \textbf{(8)}$$

Table 1. Effect comparison of the KPI clustering model

Motheds	Train time	Error
Kshape	43.839971	0.16418703
PAA-Kshape	4.974798	0.09680509
IAE-PAA-Kshape	3.767961	0.02879727

As can be seen from the above table, both the running time and accuracy of the proposed IAE-PAA-Kshape algorithm are significantly higher than the Kshape and PAA-Kshape algorithms, which also means that the KPI fast clustering model based on the IAE-PAA-Kshape algorithm proposed in this paper has a good practical effect in KPI data clustering.

5 Conclusion

The IP-Kshape clustering algorithm proposed in this paper is compared and analyzed through experiments, and it can be concluded that the efficiency and accuracy of the proposed method are significantly higher than other methods mentioned in the experiment. Using the IAE model to denoise the original KPI can better extract the timing characteristics of the KPI and reduce data complexity; using the PAA-Kshape algorithm for KPI clustering can simultaneously improve its clustering speed and accuracy. With the development of the Internet in the era of big data, Internet operation and maintenance data will gradually increase, and there will be more and more application scenarios for the intelligent operation and maintenance of KPI data. The KPI clustering involved in this article is only one of many applications. Through the clustering analysis of KPI data, further mining and analysis can be performed to complete more application scenarios such as KPI data anomaly detection or root cause analysis.

6 Acknowledgments

This work is supported by Jilin Provincial Department of Education Science and Technology Research Project (Project No: JJKH20220117KJ).

References

[1] D. Pei, S.-L. Zhang, C.-H. Pei, Machine Learning-Based Intelligent Operation and Maintenance, Communications of the CCF 13(12)(2017) 68-73.
[2] W. Huang, Research on AIOps Technologies Based on Machine Learning, [dissertation] Beijing: Beijing Jiaotong University, 2019.
[3] S.-L. Zhang, X.-F. Lin, Y.-Q. Sun, Y.-Z. Zhang, D. Pei, Research on Unsupervised KPI Anomaly Detection Based on Deep Learning, Frontiers of Data & Computing 2(3)(2020) 87-100.
[4] Y. Zhao, H. Huo, H.-T. Xu, Application of Two-Stage Random Forest Classification Method to Low-Frequency Oscillation Monitoring, Journal of Northeast Electric Power University 40 (2)(2020) 60-67.
[5] Y.-Q. Sun, Y.-J. Zhao, Y. Su, D.-P. Liu, X.-H. Nie, Y. Meng, S.-W. Cheng, D. Pei, S.-L. Zhang, X.-P. Qu, X.-Y. Guo, HotSpot: Anomaly Localization for Additive KPIs with Multi-Dimensional Attributes, IEEE Access (6)(2018) 10909-10923.
[6] Z.-Y. Li, C.-Y. Luo, Y.-W. Zhao, Y.-Q. Sun, K.-X. Sui, X.-P. Wang, D.-P. Liu, X. Jin, Q. Wang, D. Pei, Generic and Robust Localization of Multi-dimensional Root Causes, in: Proc. 2019 IEEE 30th International Symposium on Software Reliability Engineering (ISSRE), 2019.
[7] Y.-J. Wang, Y. Cao, P.-Z. Dou, F.-J. Cao, A Novel Traveling Wave Fault location Method Based on Segment Locationfor Distribution Network Hybrid Transmission Line, Journal of Northeast Electric Power University 40(4)(2020) 55-60.
[8] J. Zhou, X.-W. Wang, Y.-Y. Li, Y. Wang, J. Qiao, Location of Single Phase to Ground Fault Section in Small Current Grounding System Based on the Characteristics of Transient Zero Sequence Current Skewness, Journal of Northeast Electric Power University 40 (4)(2020) 41-47.
[9] R. Ding, Q. Wang, Y.-N. Dang, Q. Fu, H.-D. Zhang, D.-M. Zhang, Yading: fast clustering of large-scale time series data, Proceedings of the VLDB Endowment 8(5)(2015) 473-484.
[10]Z.-H. Li, Y.-J. Zhao, R. Liu, D. Pei, Robust and rapid clustering of kpis for large-scale anomaly detection, in: Proc. 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), 2018.
[11]D.-P. Kingma, M. Welling, Auto-Encoding Variational Bayes, in: Proc. 2nd International Conference on Learning Representations, ICLR 2014, 2014.
[12]X.-X. Han, J.-Z. Li, H. Gao, PAA: An Efficient Approximate Aggregation Algorithm on Massive Data, Journal of Computer Research and Development 51(1)(2014) 41-53.
[13]C.-H. Ma, X.-Q. Weng, Early Classification of Time Series Based on Piecewise Aggregate Approximation, Computer Science 45(2)(2018) 291-296+317.
[14]M. Yang, X.-N. Yu, An Ultra-Short Term Combined Prediction Considering Wind Farm Power Climbing, Journal of Northeast Electric Power University 42 (1)(2022) 63-70.
[15]W.-G. Li, H.-W. Liu, X.-Y. Liu, L.-M. Chen, L.-J. Guo, Voltage Synchronization Signal Detection Technology Based On

Improved SOGI － PLL, Journal of Northeast Electric Power University 41(3)(2021) 93-100.
[16]J. Paparrizos, L. Gravano, k-shape: Efficient and accurate clustering of time series, in: Proceedings of the 2015 ACM SIGMOD international conference on management of data, 2015.