Hindawi
Mathematical Problems in Engineering
Volume 2023, Article ID 1039287, 11 pages
https://doi.org/10.1155/2023/1039287

Research Article

Analysis of the Abnormality of Traction Energy Consumption in Urban Rail Transit System

Xinjun Gao and Xuetao Shi

Signal and Communication Research Institute of China Academy of Railway Sciences Group Co.,Ltd, Beijing 100081, China

Correspondence should be addressed to Xuetao Shi; xuetaosh@163.com

Received 16 October 2022; Revised 13 December 2022; Accepted 7 April 2023; Published 22 June 2023

Academic Editor: Li Zhu

Traction energy consumption (TEC) is a critical part of the total energy consumption in urban rail transit (URT) systems. Energy consumption patterns and abnormal analysis of TEC guarantee the energy-saving URT operation. With the rapid development of urbanization, the current energy consumption is becoming more and more prominent with some inherent drawbacks, such as complex original data, complicated statistical analysis, and abnormal energy consumption. This paper proposes a method for time accrual abnormal analysis of TEC. The system architecture of TEC typical values is presented, composed of three elements: research object, evaluation index, and time scale. The time series prediction algorithm calculates the typical values of the cumulative energy consumption index in each energy consumption mode. For the abnormality in TEC mode, the distance of the string vector is used as the similarity measure. Then, the similarity-based anomaly analysis method is used to judge the pattern abnormality. By comparing the advantages and disadvantages of engineering practice and theoretical research methods, we analyze the applicability of traditional anomaly detection algorithms to perform anomaly analysis of TEC in URT systems. The adopted time accrual abnormal analysis achieves a high fault detection rate, outperforming other models.

1. Introduction

The urban rail transit (URT) system is the main backbone of the passenger transportation system, with a large passenger capacity, short operation cycles, and low unit energy consumption. By 2022, a total of 541 cities in 79 countries have put URT systems into use, with a total mileage of more than 36,854 km. Although URT systems in China, represented by the subway system, have a short construction history, it has developed rapidly since the 21st century, as seen in Figure 1. The URT system has the advantage of energy saving in the condition of lower unit energy consumption. However, with the increasing operating mileage and passenger volume, the total energy consumption is rising. Primarily, traction energy consumption (TEC) has increased, accounting for 50% of the whole energy consumption systems. The onboard energy consumption recorder records a large amount of TEC data in the form of time accumulation. Its energy consumption mode and abnormal analysis are significant to the URT energy-saving operation.

The URT system usually uses manual meter readings and statistics to obtain the TEC data, which is prone to transcription errors and has a large workload. In recent years, some trains have recorded the data during train operation by installing sensors and metering modules, thus accumulating a large amount of energy consumption-related data mainly in discrete time series. Currently, metro still uses a fixed period to count the accumulated TEC and operating mileage data, calculate the unit consumption index per 100 vehicle kilometers, and determine whether there is a TEC anomaly by combining the threshold. However, this period is too long to detect the TEC anomaly in time. At the same time, it is not practical for data to be recorded during train operation, and the threshold is used to judge the abnormality in a one-size-fits-all manner.

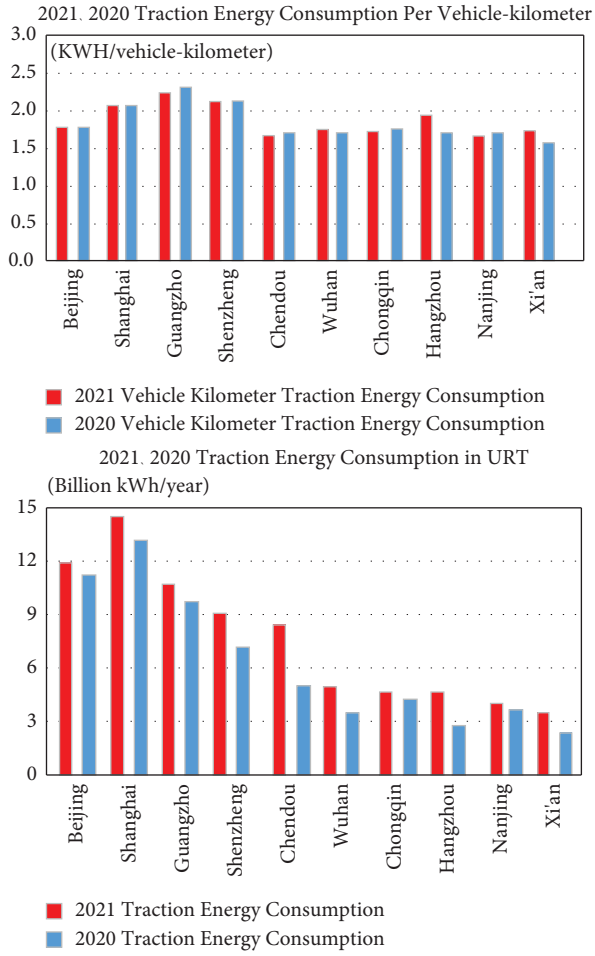

Figure 1: Development of urban rail transit in China.

There is massive data on TEC, but it fails to fully use it and clarify more targeted abnormal detection. Thus, a unified threshold is set for the abnormal analysis based on periodic statistics and empirical conclusions. By comparing and analyzing the advantages and disadvantages of engineering practice and theoretical research methods, this paper explores the applicability of traditional anomaly detection algorithms to anomaly TEC analysis. The main contributions of this work are as follows:

(i) We propose typical TEC values comprising three elements, namely, research object, evaluation index, and time scale. The evaluation index determines the specific analysis framework under different time scales.

(ii) Our method calculates typical values based on time series data, using symbolic approximate aggregation, and clustering algorithms to analyze the specific TEC patterns. The prophet algorithm predicts typical cumulative energy consumption index values for each energy consumption mode.

(iii) We use a similarity-based anomaly analysis method to detect the TEC pattern abnormalities, using string vector distance as a similarity measure.

2. Related Work

The energy consumption in the URT system caused by rapid development has attracted the attention of many scholars at home and abroad. The energy consumption analysis methods rarely use the natural energy consumption data generated in the actual train operation for analysis and mining. Even fewer scholars use these data to conduct relevant studies on abnormal TEC analysis.

By analyzing and comparing the energy consumption data during the actual train operation, Lukaszewicz divided the influencing factors of TEC into three categories, namely, basic parameters, driving strategies, and external factors [1]. Xiaobin et al. analyzed the energy consumption component of the traction system and its influencing factors by collecting TEC data [2]. Bo and Hui analyzed the influence of TEC, such as train mass, line slope type, and running resistance through simulation methods [3]. Based on the measured data, REN et al. divided the factors affecting the TEC into three categories as follows: infrastructure conditions, transportation organization, and external environment, and quantitatively analyzed the train features and lines contained in each category through data mining methods [4].

The anomaly detection approach combines many fields, e.g., machine learning and statistics. It can be divided into density-, proximity-, and model-based anomaly detection algorithms according to the anomaly detection implementation. As a typical density-based anomaly detection approach, the local outlier factor (LOF) algorithm can simultaneously determine anomalies and quantitatively analyze the anomaly degree [5]. Xiaoxia et al. made a clustering analysis on the original energy consumption data to obtain different energy consumption characteristics. The decision tree method is used to detect energy consumption modes in classified datasets. The dynamically collected data can detect anomalies based on the LOF algorithm, which can analyze the anomalies of each sampling point [6].

In proximity-based anomaly detection, anomalies are defined as objects far away from most data. Its core lies in defining the proximity of data objects, e.g., Euclidean distance, Jaccard, and cosine similarity measure. Based on MapReduce architecture, Cao et al. proposed a distance-based outlier detection (DOD) method for a distributed database system with TB-level volume. This method can realize anomaly detection under massive data with less communication cost [7]. Bin and Yifei constructed robust mean and covariance matrix estimators and proposed an anomaly detection approach based on robust Mahalanobis distance to detect outliers caused by registration errors and measurement errors [8]. Proximity-based algorithm to calculate the proximity of a large number of data has large time and space complexity, high calculation cost, and poor applicability for datasets with sizeable regional density changes.

Model-based anomaly detection requires statistical models to describe normal data and detect abnormal data. Hong proposed a new test statistic based on the sample quantile for extreme value distribution, suitable for simple data elimination tasks [9]. Habib et al. used clustering

algorithms and normalization to detect anomalies in the sensor data [10]. Peng et al. filtered noise data and used clustering algorithms to detect node anomalies in wireless sensor networks [11]. Tang Shulu et al. used density peak clustering to detect abnormal targets in low-dimensional space for hyperspectral image data [12].

3. Architecture of Specific TEC in URT System

This chapter proposes the system architecture of TEC typical values, discussing their significance and constructing a standard value system with three elements, namely, index, time scale, and specific TEC patterns and energy consumption indexes. An anomaly analysis framework from mode to point is proposed based on the standard value system.

3.1. Composition of the TEC Typical Values. The data analysis of TEC prioritizes prediction and evaluation, with little focus on anomaly detection. Anomaly detection in other fields often employs machine learning algorithms, but it lacks robust explanations and practical applicability. The improved TEC evaluation scheme expands on the existing methods, including line unit consumption, typical values, and a comprehensive evaluation index. The scheme comprises three elements as follows: research object, evaluation index, and time scale, allowing for evaluation of lines, trains, and power consumption units on hourly, daily, weekly, monthly, and yearly scales [13]. Its architecture is depicted in Figure 2.

The upgraded TEC assessment plan uses historical data as a standard for identifying TEC anomalies during regular URT operation. The energy consumption data is collected, preprocessed, and compared to the established TEC plan. Analysts identify the cause of any abnormality and guide engineers to address the issue, as shown in Figure 3.

4. TEC Anomaly Analysis Method

Line level: the accumulated TEC data collected and uploaded during all operations on a certain line and the mileage data recorded by the transportation management system (TMS) were summarized, the incorrect data were eliminated, and then the TEC data for each statistical period were calculated.

Train level: firstly, discrete time series are constructed based on the historical cumulative TEC data generated by train operation. Specific TEC patterns were obtained by mode analysis and verified by combining common operation diagrams and training sets. The first step of data analysis is to determine whether there is an abnormal TEC mode by comparing its similarity with each typical TEC pattern. Then, the accumulated energy consumption values were compared at each point of the peculiar energy consumption mode with the standard weight of the energy consumption index point by the end to determine the abnormal time point.

Energy consumption unit level: TEC is divided into traction unit energy consumption and auxiliary energy consumption. Also, the same analysis process as that of the train is adopted for traction unit energy consumption, i.e., the analysis of abnormalities from the typical mode to the specific values of the cumulative energy consumption index. For the auxiliary energy consumption, the accumulated energy consumption index per unit of time is solved, combined with the usage of additional equipment to determine the abnormality.

4.1. Anomaly Detection Algorithm

4.1.1. Data Dimensionality Reduction. The anomaly detection algorithm is as seen in Algorithm 1. First, we need to reduce the accumulated TEC dimension. Given a time series of lengths $Q = q_1$ and $q_2, \ldots, q_m$. We turn it into a data sequence of length w. $Q' = q_1', q_2', \ldots, q_w'$, where $w < m$. Then, the compression ratio for dimensionality reduction of time series data is k, and q_i' satisfies the following equation:

$$k = \frac{m}{w},$$
$$q_i' = \frac{1}{k} \sum_{j=k(i-1)+1}^{k*i} q_j, i = 1, 2, \ldots, w. \tag{1}$$

In order to reduce the dimensionality further, the Piecewise Aggregate Approximation (PAA) is usually applied prior to the symbolic aggregate approximation (SAX). SAX is used to transform a sequence of rational numbers (i.e., a time series) into a sequence of letters (i.e., a string). An illustration of a time series of 128 points converted into the word of 8 letters. Besides, we use the 4 symbol alphabets a, b, c, and d as in Figure 4. The cut lines for this alphabet are shown as the thin blue lines on the plot given below.

4.1.2. Z-Normalize Data. Before transforming time series with PAA, we Z-normalize data. Time series subsequences tend to have a high Gaussian distribution. The standardization step is based on the Z-score method, where the original dataset is transformed to satisfy the Gaussian distribution of $N(0,1)$, $\mu = 0$, and $\sigma = 1$, and the standardization formula is as follows:

$$Q' = \frac{q_i' - \mu}{\sigma}. \tag{2}$$

The normalized time series has a Gaussian distribution, which is discretized using a sequence of breakpoints denoted as B. The breakpoints partition the distribution into equal probability intervals, and sequence values are approximated using the breakpoint list and PAA. The $\mu - 1$ values of the breakpoint list correspond to the standard normal distribution random variables, as shown in Table 1. The probability values of the Gaussian distribution corresponding to the adjacent breakpoints are equal.

4.1.3. PAA Follows the Standard Procedure. To detect anomalous patterns in feature data, we convert the time series to PAA representation and then to symbols. We use a pattern discovery algorithm combined with a time series

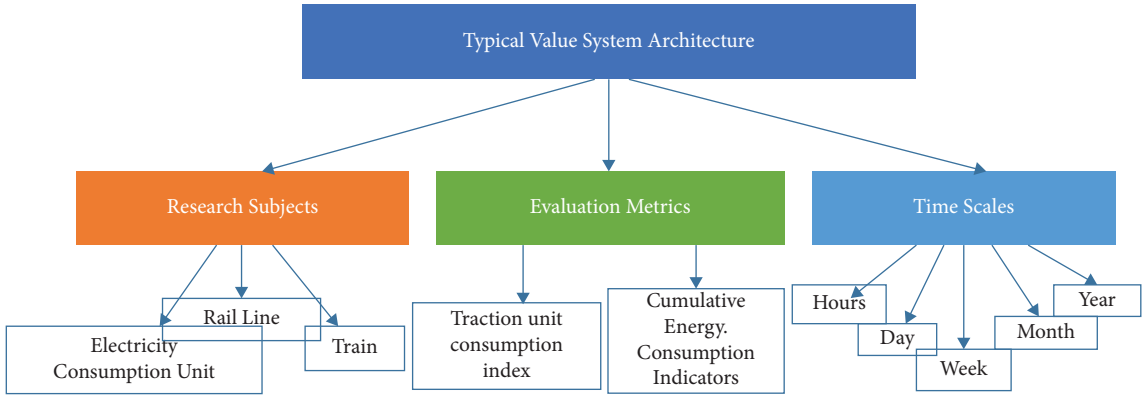

Figure 2: The structure of TEC typical values.

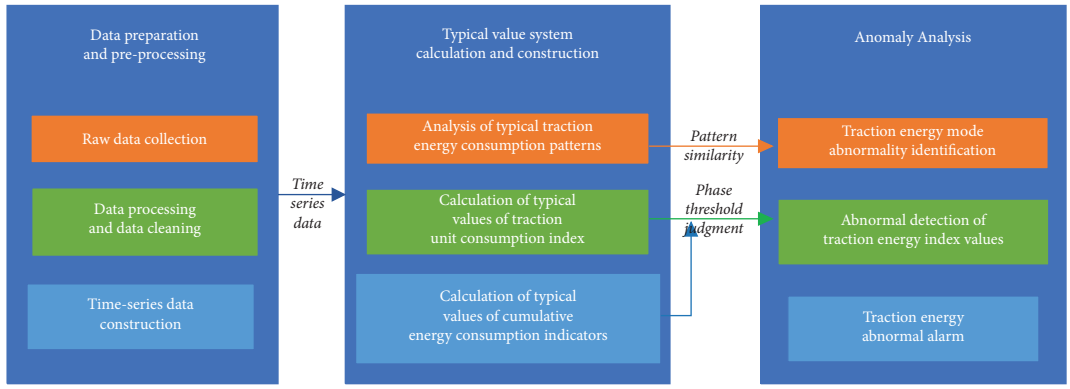

Figure 3: The framework of outlier analysis.

Require: T:24-dimensional original TEC time series data, n:the length of time series discord.
Ensure: The length of Discord and strings of length ω.

```
(1)  function Z-normalization(ts)
(2)    ts.mean←mean(ts)
(3)    ts.dev←sd(ts)
(4)    (ts − ts.mean)/ts.dev
(5)    ts_znorm = Z-NORMALIZATION (ts)
(6)    function PAA (ts, paa_size)
(7)      len == paa_size
(8)      if len == paa_size then
(9)        ts
(10)     else if len%%paa_size == 0 then
(11)       colMeans (matrix (ts, nrow = len %/% paa_size, byrow = F))
(12)     else
(13)       res = rep.int (0, paa_size)
(14)     end if
(15)     return s_paa = paa(ts_znorm, paa_size)
(16) end function
(17) Use the 4 symbols alphabet a,b,c,d
(18) SAX transform of ts into string through 9-points PAA: "baabccbc":
(19) ts_to_string (dat_paa_9, cuts_for_asize (9))
(20) discords = find_discords_hotsax(dd)
```

Algorithm 1: Discretization of the PAA representation of time series into SAX.

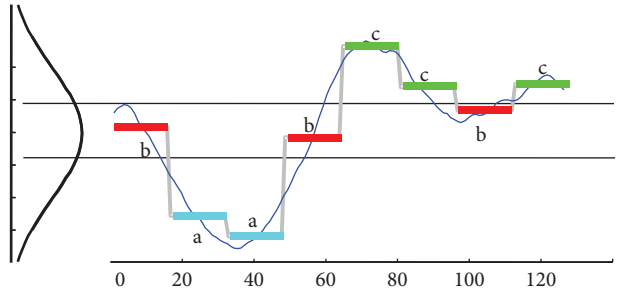

Figure 4: Time series processed by PAA.

Table 1: The breakpoints.

	μ						
β_i	3	4	5	6	7	8	9
β_1	−0.43	−0.67	−0.84	−0.97	−1.07	−1.15	−1.22
β_2	0.43	0	−0.25	−0.43	−0.57	−0.67	−0.76
β_3		0.67	0.25	0	−0.18	−0.32	−0.43
β_4			0.84	0.43	0.18	0	0.14
β_5				0.97	0.57	0.32	0.14
β_6					1.07	0.67	0.43
β_7						1.15	0.76
β_8							1.22

distance metric based on the nonlinear statistical feature representation. However, relying only on the mean can result in the lost information, as two series with different patterns can have the same mean and variance. Therefore, we also use morphological features like slope and angle to accurately represent a time series. The slope values for each segment of the compressed subsequence can be calculated using the following equations:

$$\overline{q}_i = \frac{k\sum_{j=j_0}^{k*i} jq_j - \left(\sum_{j=j_0}^{k*i} j\right)\left(\sum_{j=j_0}^{k*i} q_j\right)}{k\sum_{j=j_0}^{k*i} j^2 - \left(\sum_{j=j_0}^{k*i} j\right)^2}, \tag{3}$$

$$j_0 = k(i-1) + 1. \tag{4}$$

5. Experiments

This section uses the actual operating energy consumption data collected from a train set of the Beijing URT Operating Company. The train consisted of 6 motor vehicles without cabs and 2 trailer vehicles with cabs, and data collection involved recording second-level cumulative traction unit energy consumption for each motor vehicle and secondary cumulative auxiliary energy consumption for each trailer. Hourly cumulative TEC data for the train's daily traction energy mode analysis was obtained by processing the original data file. The resulting 24 data points form a 24-dimensional original TEC time series data $Q = q^1, q^2, \ldots, q^{24}$, with each dimension representing the cumulative TEC of each period.

The data are divided into 24 dimensions, and a morphological feature-based symbolic representation method is used to identify three TEC patterns in the time series data. The algorithm involves transforming the parameter time series into characters with actual semantics by first converting the original time series into the PAA representation and then converting the PAA data into a string. The algorithm is implemented in Python and is available on PyPi for installation using pip.

5.1. Characteristics of TEC Data. This study utilizes data from the energy consumption metering devices installed on several lines and train groups in the Beijing Subway. The device records instantaneous voltage and current values and accumulated energy consumption, consisting of a voltage sensor, current sensor, and metering module per vehicle [14]. Each vehicle is equipped with a set of multitrain energy consumption measurement devices, as depicted in Figure 5.

Each metering device wirelessly uploads data to the database terminal, which can be analyzed to retrieve instantaneous voltage, current, and accumulated energy consumption values. The motor train data file records the cumulative energy consumption and regenerative energy of the traction unit, while the trailer data file records the cumulative energy consumption of all auxiliary equipment powered by the additional inverter. Tables 2 and 3 present the data for the motor train and trailer, respectively.

The original data file records regenerative energy as negative due to electric braking, and trains frequently switch between traction and braking, resulting in fluctuating instantaneous voltage and current values. Analyzing energy consumption based on voltage and current values alone is challenging, so this study uses the accumulated energy consumption as the research object. Hourly cumulative energy consumption values are obtained from the original data file, and discrete univariate time series data is

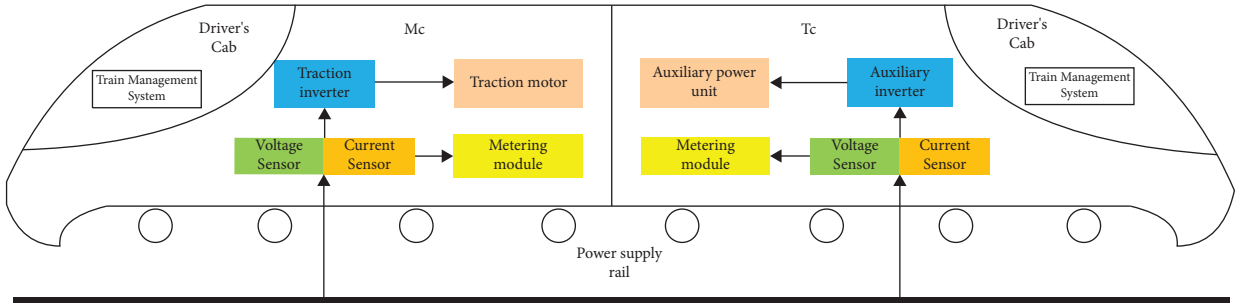

Figure 5: The structure of metering device.

Table 2: Dataset of motor train.

Time	Voltage (V)	Electric current (A)	Cumulative energy consumption (kW·h)	Regenerative energy (kW·h)
2022/5/21 11:34:01	813.3	380.4	248.1	−11.0
2022/5/21 11:34:02	838.9	268.4	248.2	−11.0
2022/5/21 11:34:03	844.9	219.3	248.3	−11.0
2022/5/21 11:34:04	857.7	152.0	248.3	−11.0
2022/5/21 11:34:05	866.2	93.9	248.3	−11.0
2022/5/21 11:34:06	873.0	56.1	248.4	−11.0

Table 3: Dataset of trailer train.

Time	Voltage (V)	Electric current (A)	Cumulative energy consumption (kW·h)
2022/5/21 11:34:01	886.8	22.5	109.5
2022/5/21 11:34:02	886.4	23.4	109.5
2022/5/21 11:34:03	887.0	23.8	109.5
2022/5/21 11:34:04	886.4	23.7	109.5
2022/5/21 11:34:05	886.6	24.0	109.6
2022/5/21 11:34:06	887.0	22.9	109.6

constructed based on these values. The time series data consists of cumulative energy consumption values from a train on the Beijing Subway between August 5th and August 11th, 2021. The time series curve of energy consumption is depicted in Figure 6.

The TEC level is impacted by the complex energy flow process and changeable train operating conditions. In engineering practice, the TEC index's general value for each line is determined based on the historical statistical data, and the fluctuation interval is set as the threshold for rough abnormal judgment. Figure 7 shows the average traction unit consumption of the Beijing metro based on the field investigation and historical data statistics.

TEC time series data display periodic and seasonal characteristics, with energy consumption values affected by the total load rate and auxiliary equipment opening. Fluctuations in the time series curve are complex and typically contain multiple peaks. The trend of the time series represents changes in train TEC levels over time, with each point's energy consumption value related to the adjacent periods. The original dataset used in this study is the data file uploaded by the energy consumption metering device, with high data quality and few errors or missing data despite occasional failures in collection, calculation, storage, transmission, and analysis links. Time series subsequences tend to have a high Gaussian distribution in Figure 8.

5.2. Similarity Measure. The similarity measure, as a measure of how close two things are, is used to measure the anomalies in a single time series, as shown in Figure 9. The closer two things are, the more similar they are, while the farther away two things are, the less similar they are. Dist is a function that takes sequences Q ($Q = q_1 \ldots q_m$) and C ($C = c_1 \ldots c_m$) as parameters and returns a non-negative value R, which is considered as the similar distance between the two and must be symmetric. Their Euclidean distance is defined as the first equation given below. In the second equation, PAA distance lower-bounds the Euclidean Distance.

$$D(Q,C) \equiv \sqrt{\sum_{i=1}^{m} (q_i - c_i)^2}, \tag{5}$$

$$D_{R(\overline{Q},\overline{C})} \equiv \sqrt{k}\sqrt{\sum_{i=1}^{w} (\overline{q}_i - \overline{c}_i)^2}. \tag{6}$$

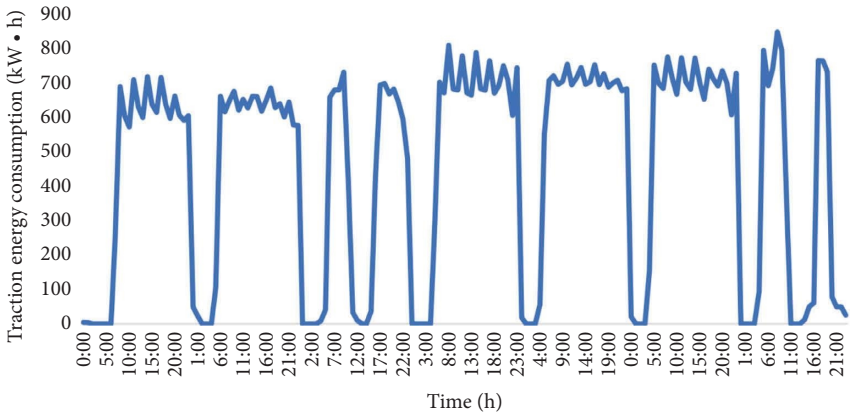

Figure 6: The time series curve of energy consumption.

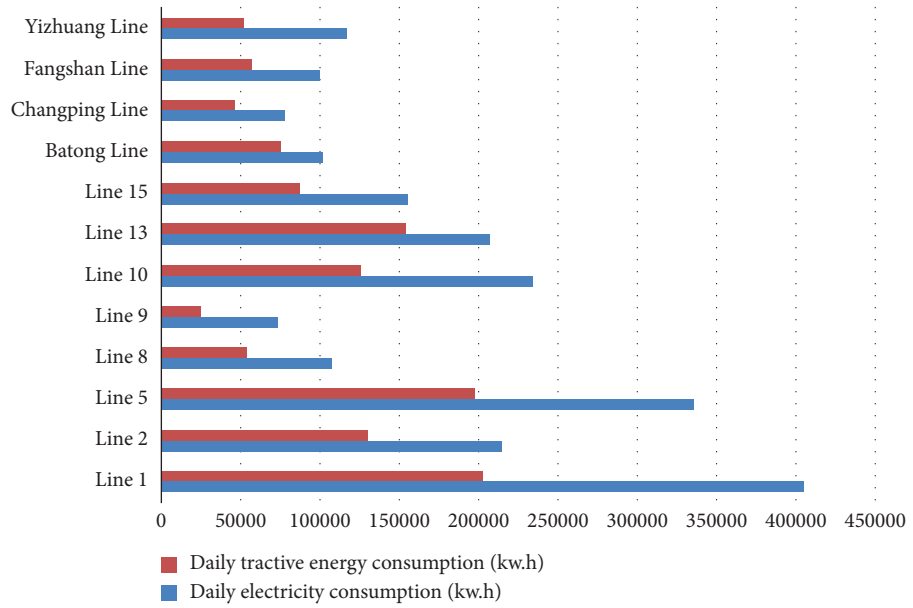

Figure 7: The average traction unit consumption of each line in Beijing subway.

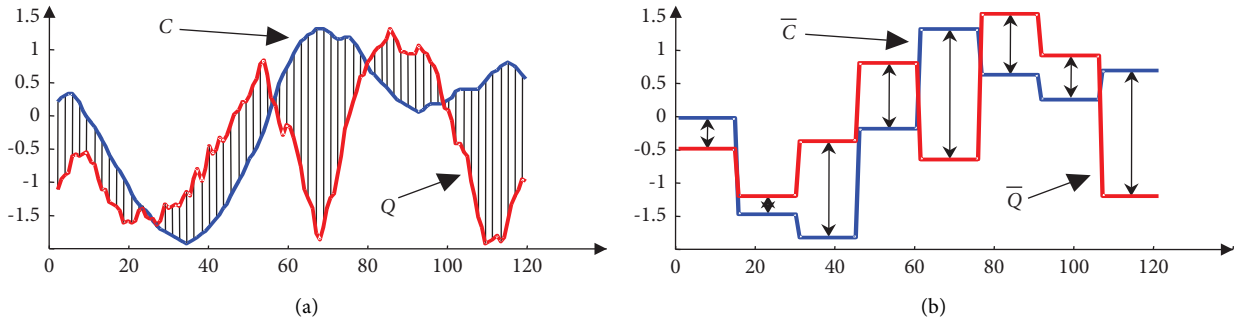

Figure 8: (a) PAA distance lower-bounds the Euclidean distance. (b) Euclidean distance.

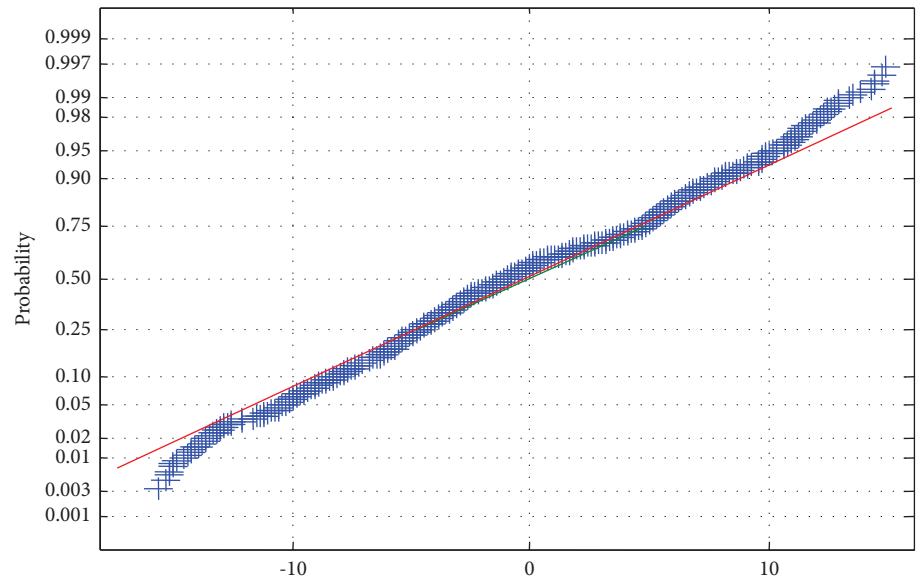

FIGURE 9: A normal probability plot of the distribution of values from subsequences of length 128.

Equation (7) defines a function that computes the minimum distance between the string representations of the original time series O and C. This function can be efficiently implemented using table lookup. Additionally, time series subsequences exhibit a Gaussian distribution, which is a characteristic tendency.

$$\mathrm{MTNDIST}(\widehat{Q},\widehat{C}) \equiv \sqrt{k}\sqrt{\sum_{i=1}^{w}\left(\mathrm{dist}\left(\widehat{q}_i,\widehat{c}_i\right)\right)^2}, \tag{7}$$

where the dist function is implemented by using the lookup table for the particular set of the breakpoints (alphabet size), as shown in the table below, and where the singular values for each cell (q, c) is computed as follows:

$$\mathrm{cell}(q,c) = \begin{cases} 0, & \text{if} |q-c| \le 1, \\ \beta_{\max(q,c)-1} - \beta_{\min(q,c)-1}, & \text{otherwise.} \end{cases} \tag{8}$$

To convert a time series of an arbitrary length to SAX, we need to define the alphabet cuts. Saxpy retrieves cuts for a normal alphabet (we use size 3 here) via cuts_for_asize function: *from saxpy. alphabet import cuts_for_asize*

Cuts_for_asize(3).

First, we use the "ts_to_string" function to convert a time series into letters using SAX. However, before applying this function, we must z-normalize the input time series using a normal alphabet to obtain a string, such as abcba. Next, to investigate the structure of the input time series and identify any anomalous (i.e., discords) or recurrent (i.e., motifs) patterns, we used the "time series to SAX conversion via sliding window" approach. This approach is commonly employed, and Saxpy implements this workflow. The result is represented as a data structure of resulting words and their respective positions on time series as follows:

defaultdict (list,
'aac' : [4, 10, 11, 30, 35],
'abc' : [12, 14, 36, 44],
'acb' : [5, 16, 21, 37, 43],
'acc' : [13, 52, 53],
'bac' : [3, 19, 34, 45, 51],
'bba' : [31],
'bbb' : [15, 18, 20, 22, 25, 26, 27, 28, 29, 41, 42, 46],
'bbc' : [2],
'bca' : [6, 17, 32, 38, 47, 48],
'caa' : [8, 23, 24, 40],
'cab' : [9, 50],
'cba' : [7, 39, 49],
'cbb' : [33],
'cca' : [0, 1])

Anomalies in TEC patterns are defined as operating conditions that deviate significantly from the specific TEC patterns. In time series data mining, retrieval, clustering, classification, summary, and anomaly detection are usually performed based on series similarity, including temporal similarity, shape similarity, and change similarity. Similarity measures based on Minkowski distance, cosine similarity, correlation, and mutual information are often used to measure the similarity of two time series. Euclidean distance model is simple, intuitive, easy to understand, and fast, and it is often used to measure the similarity of discrete time series.

The Euclidean distance between the two time series can be expressed as the square root of the sum of the squared differences of each pair of corresponding points. The distance metric defined by the PAA approximation can be viewed as the square root of the sum of the squared differences between each pair of corresponding PAA coefficients multiplied by the square root of the compression rate, as shown in Figure 10. The distance between two SAX representations of a time series requires finding the distance

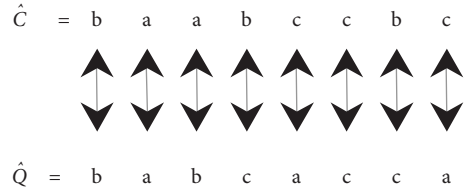

Figure 10: The symbols between two time series after PAA.

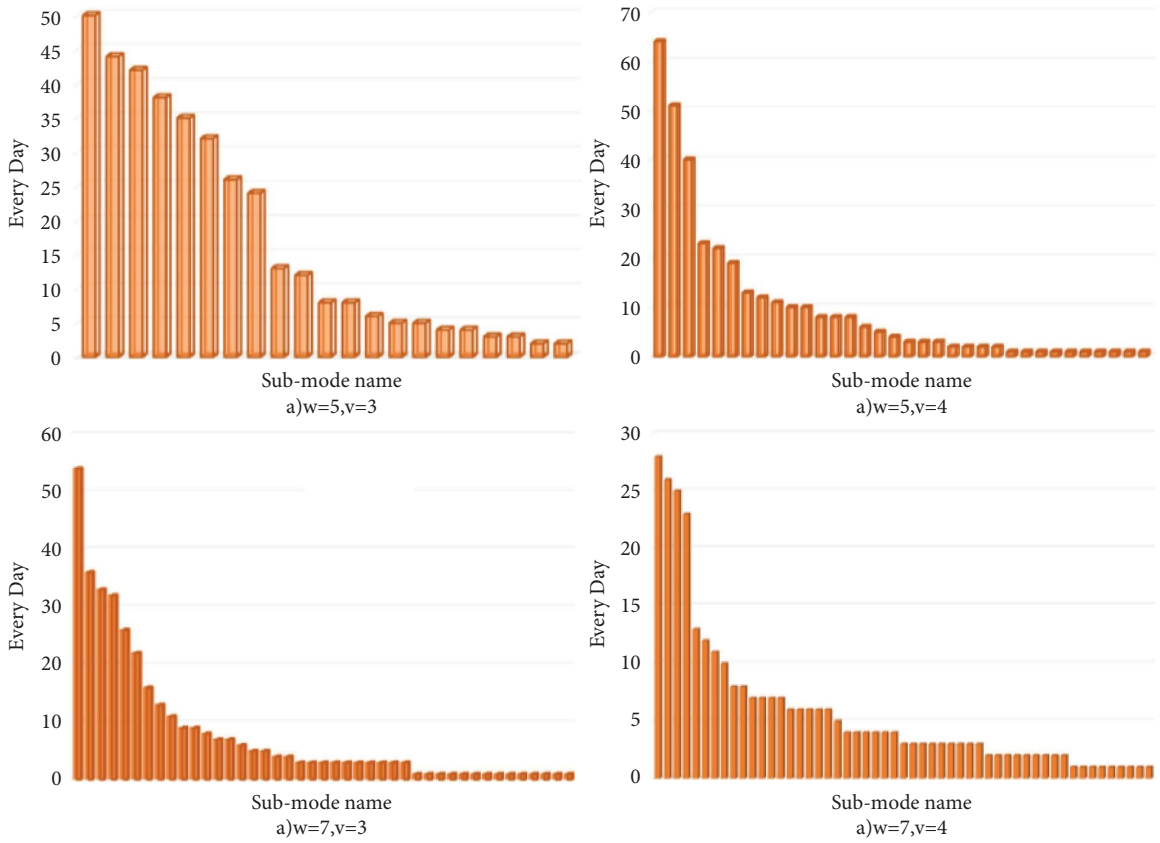

Figure 11: Number of patterns and days under different values.

between each pair of symbols, squaring them, summing them, taking the square root, and finally multiplying by the square root of the compression rate. By rigorous proof, we get it in the following equation:

$$\sqrt{\sum_{i=1}^{n}(q_i - c_i)^2} \geq \sqrt{\frac{n}{w}}\sqrt{\sum_{i=1}^{w}(\overline{q_i} - \overline{c_i})^2 n(\overline{Q} - \overline{C})^2} \geq n(\mathrm{dist}(\hat{Q}, \hat{C}))^2. \quad (9)$$

5.3. Experimental Results. Higher values of ω and μ result in more detailed energy consumption levels and more complex TEC submodes, which can have a significant impact on the subsequent analysis using ML algorithms. The TEC pattern analysis aims to investigate the TEC level's variation over time within a day, using daily TEC time series data as the research object. Low energy consumption levels are represented by the letter *a*, high energy consumption levels by *c*, and medium energy consumption levels by *b*. For example, on August 11th, 2021, when $\omega = 7$ and $\mu = 3$, the original time series is represented as the string baabccbc after processing and conversion. The data of the subject train's 354 days of operation in 2021 are processed to form 42 string vectors representing various TEC variation patterns, as shown in Figure 11. For the submodel conclusions, refer to Table 4.

Table 4: Subpatterns of PAA processing.

PAA	Days
Aaabbbccc	63
Abbbbbcca	43
Aaabccccc	38
Aacbbbccc	36
Abcbbbccc	31
Ccabbbcac	27
Acabbbccc	22
Acccbbccc	19
Caabbbccc	16
Aaabcbccc	14
Aabbabccc	13
Acbbbbccc	10
Bbabbbcaa	9
Accaabccc	9
Aabbabccc	8
Abbaabccc	7
Aaaccbccc	6
Bbaaabcaa	6
Abaaabccb	5
Baababccc	5
Ccaaabccc	4
Abbaabccc	4
Aaccabccc	4
Abbaabcaa	4
Baccabccc	4
Aaaaabccc	3
Abbcabccc	3
Caaacbccc	3
Baabccbc	2
Ccabccbc	2
Baabcacc	2
Ccbbccbc	1
Acabccbc	1
Bacbbabc	1
Aaabccbc	1
Acabccbc	1
Baccbcbc	1
Acabcccc	1
Ccbaacbc	1
Caababbc	1
Cbabccab	1
Cbbaacbc	1

6. Conclusion

Based on the improved TEC evaluation scheme and anomaly analysis framework, we propose an anomaly analysis method of TEC for urban rail lines, trains, and traction units. The value anomaly detection method based on the improved TEC evaluation scheme combines the advantages of mathematical statistics, prediction algorithms, and manual experience in setting thresholds. It has a good adaptability to the characteristics of TEC data, analysis needs, and practical applications. In the numerical simulation experiments, the effectiveness of the new method for TEC analysis is verified by comparing the feature identification and anomaly detection results. Meanwhile, compared with the traditional way, the new approach is able to find and detect the anomalous patterns better and has stronger robustness.

Data Availability

No datasets are available to support the findings of this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper is supported by the Beijing Natural Science Foundation (L221016).

References

[1] P. Lukaszewicz, *Energy consumption and running times for trains issn 1103-470x; isrn kth/fkt/d-01/25-se*, SAGE, Tyne, UK, PhD, 2001.

[2] Z. Xiaobin, B. Yun, and Z. Shanshan, "Metro train energy consumption characteristics based on empirical data analysis," *Journal of Transportation Systems Engineering and Information Technology*, vol. 21, no. 6, pp. 264–271, 2021.

[3] L. Bo and L. Hui, "Research on the Beijing subway energy management platform," *Information Security and Technology*, vol. 7, no. 1, 2016, (in Chinese).

[4] J. H. Ren, Q. Zhang, and F. Liu, "Analysis of factors affecting traction energy consumption of electric multiple unit trains based on data mining," *Journal of Cleaner Production*, vol. 262, no. 1, pp. 47–58, 2020.

[5] J. Hang, L. Tun, G. Hansu, D. Xianghua, and G. Ning, "A dynamic real-time abnormal detection method for university building energy consumption," *Computer Engineering*, vol. 34, no. 4, p. 7, 2017.

[6] Q. Xiaoxia, X. Dan, and W. Bo, "Data mining method for real-time monitoring of energy consumption," *Journal of Chongqing University*, vol. 35, no. 7, pp. 133–137, 2012.

[7] L. Cao, Y. Yan, C. Kuhlman, Q. Wang, and E. A. Rundensteiner, "Multi-tactic distance-based outlier detection," in *Proceedings of the IEEE International Conference on Data Engineering*, San Diego, CA, USA, April 2017.

[8] W. Bin and C. Yifei, "Multivariate outlier detection based on robust mahalanobis distance," *Statistics and Decision Making*, vol. 3, 2005.

[9] Z. Hong, "Test of multiple outliers in type i extreme distribution samples," *Journal of UEST of China*, vol. 18, 2008.

[10] U. Habib, G. Zucker, M. Blochle, F. Judex, and J. Haase, "Outliers detection method using clustering in buildings data," in *Proceedings of the Conference of the IEEE Industrial Electronics Society*, Yokohama, Japan, November 2015.

[11] Z. Peng, F. Xin, and Z. Jianguo, "Clustering anomaly detection algorithm based on spatial correlation in wireless sensor networks," *Application Research of Computers*, vol. 30, no. 5, p. 4, 2013.

[12] T. Shulu, Z. Chunhui, and C. Ying, *Improved Detection of Hyperspectral Anomalies Based on Density Peak Background Purification*, Engineering Journal of Heilongjiang University, Harbin, China, 2021.

[13] L. Zhu, Y. Li, F. R. Yu, B. Ning, and X. Wang, "Cross-layer defense methods for jamming-resistant cbtc systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–13, 2020.

[14] Y. Li, L. Zhu, H. Wang, F. R. Yu, and S. Liu, "A cross-layer defense scheme for edge intelligence-enabled cbtc systems against mitm attacks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–13, 2020.