An Imbalanced Sequence Feature Extraction Approach for the Detection of LTE-R Cells with Degraded Communication Performance

Jiantao Qu 1,2,* , Chunyu Qi 2 and He Meng 2

1 School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
2 National Engineering Research Center for Digital Construction and Evaluation Technology of Urban Rail Transit, China Railway Design Corporation, Co., Ltd., Tianjin 300308, China; qichunyu@crdc.com (C.Q.); 16112083@bjtu.edu.cn (H.M.)
* Correspondence: 16112084@bjtu.edu.cn

Abstract: Within the Shuo Huang Railway Company (Suning, China) the long-term evolution for railways (LTE-R) network carries core wireless communication services for trains. The communication performance of LTE-R cells directly affects the operational safety of the trains. Therefore, this paper proposes a novel detection method for LTE-R cells with degraded communication performance. Considering that the number of LTE-R cells with degraded communication performance and that of normal cells are extremely imbalanced and that the communication performance indicator data for each cell are sequence data, we propose a feature extraction neural network structure for imbalanced sequences, based on shapelet transformation and a convolutional neural network (CNN). Then, to train the network, we set the optimization objective based on the Fisher criterion. Finally, using a two-stage training method, we obtain a neural network model that can distinguish LTE-R cells with degraded communication performance from normal cells at the feature level. Experiments on a real-world dataset show that the proposed method can realize the accurate detection of LTE-R cells with degraded communication performance and has high practical application value.

Keywords: LTE-R; abnormal detection; shapelet transformation; imbalanced data classification; feature extraction

1. Introduction

In recent years, due to the rapid development of the Chinese economy, demand for the transportation of goods has also rapidly increased, bringing great challenges to China’s transportation industry. In this situation, heavy haul railways have become a main development direction for freight railways [1]. Heavy haul trains have bigger load-carrying capacities and longer lengths, making them more difficult to operate. The existing wireless communication system, based on the global system of mobile communications for railways (GSM-R), has the disadvantages of insufficient bandwidth, poor anti-interference ability, and insufficient security [2]. Given this, Shuo Huang Railway applies the long-term evolution for railways (LTE-R) network to carry core wireless communication services, such as the synchronous control and wireless reconnection of heavy haul trains.

The LTE-R wireless communication system carries core communication services for heavy haul trains. Therefore, the communication performance of LTE-R cells directly affects the operational safety of heavy haul trains. At present, Shuo Huang Railway mainly collects communication performance indicator data for each cell using drive tests and then manually screens for cells with degraded communication performance by observing the deterioration of key performance indicators. However, due to the large amount of drive test data, this traditional manual analysis detection method is inefficient and accuracy largely depends on the personal abilities and work status of the operational and maintenance
personnel [3]. In fact, it is difficult to accurately determine whether a certain cell has experienced communication performance degradation solely based on road test data. The evolved radio access bearer (E-RAB) abnormal release rate (EARR) of LTE-R core communication services is a more practical indicator for evaluating the communication performance of LTE-R cells [2]. But EARR is a statistical indicator that is difficult to obtain in real time. Therefore, it is necessary to develop more efficient and intelligent methods for the detection of LTE-R cells with degraded communication performance by establishing the accurate mapping relationship between the drive test data and EARR.

Some experts and scholars have already studied detection methods for abnormal wireless communication performance. These detection methods are mainly divided into two types: shallow model methods [4–7] and deep model methods [8–12]. Shallow model methods utilize traditional shallow model machine learning methods to detect abnormal wireless communication performance. However, shallow model methods have shortcomings in extracting complex nonlinear patterns from data. With the continuous development of deep learning technology, more researchers are attempting to use deep model methods to detect abnormal wireless communication performance. Since the mapping relationship between drive test data and EARR is a complex nonlinear mapping relationship, and deep model methods have advantages in dealing with the problem of complex nonlinear mapping [13], this paper chooses to use deep model methods to establish this mapping relationship. Furthermore, due to the high reliability of the LTE-R communication system, records for cells with degraded communication performance are very rare (the ratio of normal data to abnormal data is close to 30:1), so the main difficulty in detecting LTE-R cells with degraded communication performance is how to build an imbalanced data classification model with high classification accuracy.

In order to solve the problem of detecting LTE-R cells with degraded communication performance, we fully consider the characteristic of communication performance indicator data being sequentially distributed along railways, and we treat the detection problem as an imbalanced sequence data classification problem. Then, based on the basic principle of shapelet transformation, we realize the feature extraction of sequence data by transforming CNN. In addition, to solve the problem of imbalanced data classification, we set the loss function based on the Fisher criterion and propose a two-stage training method based on contrastive loss and the differential evolution algorithm. Experiments on a real-world dataset showed that the proposed method can accurately detect LTE-R cells with degraded communication performance, which could provide an effective technological means for improving the efficiency of LTE-R network operation and maintenance and ensuring the safety of railway operations.

The main contributions of this paper can be highlighted as follows.

- Based on the principle of shapelet transformation, this paper designed a neural network structure that can extract morphological features of sequence data by transforming CNN neural networks.
- In the detection of LTE-R cells with degraded communication performance, considering the class imbalanced problem, a two-stage training method is proposed to make the features extracted by the trained feature extraction network meet Fisher criterion as much as possible.
- By using machine learning methods, the mapping relationship between the drive test data and the abnormal release rate of LTE-R core communication services was established, which provides a powerful tool for the operation and maintenance of LTE-R network.

The rest of the paper is organized as follows. Section 2 investigates related work. The preliminaries are provided in Section 3. Section 4 introduces the proposed method. Section 5 presents and discusses the experimental results. Conclusions are drawn in Section 6.
2. Related Work

2.1. Abnormal Wireless Communication Performance Detection

Accurate detection of abnormal wireless communication performance is the foundation of wireless network operation and maintenance. Presently, many detection methods have been proposed to solve this problem, which can be divided into two types: shallow model methods [4–7] and deep model methods [8–12]. For example, Chernogorov et al. [4] analyzed the event sequences reported by mobile terminals to base stations and utilized the N-gram analysis method to reduce dimensionality and classify the data, ultimately achieving the detection of abnormal wireless communication performance. Chernov et al. [5] also analyzed the event sequences reported by mobile terminals to base stations and compared the accuracy of multiple anomaly detection methods for detecting abnormal wireless communication performance caused by random access channel anomalies. Miao et al. [6] achieved high detection accuracy by using the kernel density-based local outlier factor (LOF) algorithm to detect cells with degraded communication performance. Safaei et al. [7] designed a decentralized abnormal wireless communication performance detection method based on the local outlier factor algorithm to perform abnormal detection in wireless sensor networks (WSNs). The above studies all used traditional shallow model machine learning methods to detect abnormal wireless communication performance. However, the development of neural networks, especially deep learning technologies, has provided a new research direction for the detection of abnormal wireless communication performance. Premkumar et al. [8] proposed a deep learning-based lightweight distributed denial of service attack detection scheme to detect the anomaly in the data forwarding phase of WSN. Regin et al. [9] compared the performance of the convex hull algorithm, naive Bayes algorithm, and CNN in the detection of abnormal nodes in WSN, and proved that CNN has the best detection performance. Kuadey et al. [10] proposed a framework based on a long short-term memory (LSTM) deep learning technique that detects DDoS attacks in a 5G network. Huan et al. [14] proposed an unsupervised spectrum anomaly detection method based on the variational autoencoder (VAE) and aimed to detect anomalies in unauthorized frequency bands in a wireless communication network. Dahiya et al. [12] proposed a unique DDoS attack detection model, which combined the LSTM with the recurrent neural network (RNN) to construct the detection model, and utilized the opposition learning-based seagull optimization algorithm (OLSOA) to optimize the weight of the RNN for more accurate detection of DDoS attacks. Qu et al. [3] proposed a dual encoder denoising autoencoder (DEDAE) neural network based on the non-dominated sorting genetic algorithm-III (NSGA-III) and generative adversarial network (GAN), which can detect LTE-R cells with degraded wireless communication performance.

Since wireless communication network performance degradation is the result of many factors [2], the detection of abnormal wireless communication performance is usually a complex nonlinear problem. This paper chooses the deep model method to perform the detection of LTE-R cells with degraded communication performance.

2.2. Imbalanced Data Classification

Real-world datasets are mostly imbalanced and the lack of minority class samples can cause bias in classification models that have been constructed using classification methods [15]. Currently, imbalanced data classification methods can be divided into two groups: data-level and algorithm-level methods. There are many methods for solving imbalanced data classification problems at the data level, which can also mainly be divided into two types: undersampling [16–18] and oversampling [19–21]. However, although data-level methods can effectively deal with the imbalanced data problem, when data are extremely imbalanced, the excessive generation or reduction of data can alter the distribution of the original data while the limited generation or reduction of data cannot change the problem of data imbalance. Due to the high reliability of the LTE-R wireless communication system, datasets for the detection of LTE-R cells with degraded communication performance are extremely imbalanced. Therefore, algorithm-level methods should be
considered for realizing the detection of cells with degraded performance. The main idea of algorithm-level methods is to directly modify the learning process of classification methods to improve sensitivity to minority classes. Recently, shallow-model algorithm-level methods have achieved great success in the fields of imbalanced data classification and fault detection [22–24]. However, they have significant limitations in processing high-dimensional data and expressing complex nonlinear patterns. The emergence of neural networks and deep learning technologies has provided new and effective ways to solve imbalanced data classification problems, and many new solutions for imbalanced data classification have emerged. Zhang et al. [25] applied the cost-sensitive loss function to a deep belief network (DBN) and utilized the evolutionary algorithm to find the optimal misclassification cost, thus realizing the classification of imbalanced data. Geng et al. [26] proposed a cost-sensitive deep convolutional neural network (CNN) and experimentally demonstrated the superiority of this method in imbalanced time series classification. The above cost-sensitive method involves strengthening the recognition ability of the model to minority class samples by introducing the cost-sensitive factor in the training process. However, cost-sensitive factors are often difficult to determine, and these kinds of empirical risk minimization methods often have limited effectiveness when data are extremely imbalanced. Therefore, when the data are extremely imbalanced, some researchers utilize autoencoder (AE) neural networks or generative adversarial networks (GANs) to build data models for majority class samples. They then determine whether the data are abnormal based on the reconstruction error. Jinwon et al. [27] discussed how to use AE and the variational autoencoder (VAE) neural network to detect abnormal data. Thomas et al. [28] utilized GAN to guide the training process of AE, which can help AE learn the data distribution of majority class data. Cheng et al. [29] proposed an improved autoencoder for unsupervised anomaly detection, which optimizes the anomaly detection task by integrating anomaly detection-based loss and autoencoder reconstruction loss. Furthermore, this method can learn representations that preserve local data structures to avoid feature distortion. But, these unsupervised anomaly detection methods will overlook the information contained in minority class samples. Given the above, constructing specific optimization objective functions to increase the distance between samples of different categories becomes an effective solution. Wang et al. [30] designed a change detection algorithm based on Siamese networks for optical aerial images, and utilized an improved contrastive loss–focal contrastive loss (CFCL) to train the model, which can solve the data imbalanced problem. Jiao et al. [31] proposed an autonomous anomaly detection technique for multivariate time series data based on a novel self-supervised contrastive loss, which outperforms state-of-the-art anomaly detection approaches and exhibits robustness when training data are contaminated.

The above deep learning-based imbalanced data classification methods have achieved great success in different fields. But, to realize the accurate detection of LTE-R cells with degraded communication performance, it is still necessary to study the corresponding neural network structure and training method based on the characteristics of imbalanced sequence data.

3. Preliminaries
3.1. Data Description

In this study, we chose relevant data for the operation and maintenance of the Shuo Huang Railway LTE-R network to construct a dataset for detecting cells with degraded communication performance. There are two main sources of Shuo Huang Railway operation and maintenance data: drive test data and abnormal release data from the E-RAB for each wireless communication service exported from the LTE-R Operation and Maintenance Center (OMC). Drive tests are the most commonly used methods for measuring the wireless network communication status along the Shuo Huang Railway. They are usually carried out by operation and maintenance personnel, who use measurement equipment to collect wireless communication performance indicator data from the user equipment.
(UE) side along the Shuo Huang Railway. At present, the OMC used in the Shuo Huang Railway LTE-R system is the Huawei u2000 (Huawei, Shenzhen, China), which can record historical key performance indicator data from the LTE-R system, thereby achieving unified management and maintenance of the network element layer and the network layer of the LTE-R network [32]. Figure 1 shows the source of Shuo Huang Railway operation and maintenance data.

![Figure 1](image-url)

**Figure 1.** The source of Shuo Huang Railway operation and maintenance data.

Drive test data contain a lot of information, including the reference signal receiving power (RSRP), reference signal receiving quality (RSRQ), signal-to-interference-plus-noise ratio (SINR), geographic location information, physical-layer cell identity (PCI), etc. Among these metrics, RSRP and SINR measure the communication status of LTE-R cells from the perspectives of signal strength and noise, respectively, while RSRQ is an important indicator for cell handover and reselection. Therefore, we selected RSRP, RSRQ, and SINR as the metrics for judging whether communication performance deterioration occurs in LTE-R cells. Furthermore, PCI was selected to categorize drive test data according to the LTE-R cells they belong to.

Using data exported from the OMC system, many statistical values related to LTE-R communication performance can be obtained, such as throughput, latency, dropout rate, etc. In this paper, we focused on the E-RAB abnormal release rate. The successful establishment of an E-RAB shows that eNodeB has allocated wireless resources and radio bearers to the UE. Therefore, the E-RAB abnormal release rate represents the ability of eNodeB or a cell to carry services and can be used to evaluate the operational status of eNodeB or cells [2]. In the Shuo Huang Railway LTE-R system, the E-RAB abnormal release rate is calculated separately, according to the different wireless communication service types. The different types of wireless communication services on the Shuo Huang Railway are assigned a quality-of-service (QoS) class identifier (QCI) value, ranging from 1 to 7. The different QCI values and their corresponding communication services for the Shuo Huang Railway LTE-R system are shown in Table 1.

From Table 1, it can be seen that communication services with a QCI of 1 have the highest QoS priority. The higher the QoS priority, the more important the communication service and the more resources it can obtain from wireless communication resource allocation. Therefore, we chose to use the E-RAB abnormal release rate (EARR) of communication services with a QCI of 1 as the only indicator to evaluate whether cells had experienced communication performance degradation. If the E-RAB abnormal release rate was greater than a certain threshold, the corresponding cell was considered to have degraded communication performance.
Table 1. QCI value and its corresponding services.

<table>
<thead>
<tr>
<th>Terminal Type</th>
<th>Communication Service</th>
<th>QCI Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminal of train operation control</td>
<td>Train control data transmission service</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Emergency call</td>
<td>1</td>
</tr>
<tr>
<td>The cab-integrated radio communication equipment</td>
<td>Transmission dispatching order</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Wireless train number calibration</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>In-vehicle speech</td>
<td>2</td>
</tr>
<tr>
<td>Handheld mobile station</td>
<td>Emergency call</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Speech call</td>
<td>3</td>
</tr>
<tr>
<td>Train video surveillance system</td>
<td>Video Surveillance</td>
<td>4</td>
</tr>
</tbody>
</table>

In this paper, the drive test data of each LTE-R cell are taken as feature values and the linear interpolation method is used to interpolate the drive test data into sequences of the same length. Then, based on the EARR of the communication services with a QCI of 1 exported by the OMC, the drive test data are labeled as abnormal or normal, and a dataset is constructed for the detection of LTE-R cells with degraded communication performance. Figure 2 shows the construction process for this dataset.

As can be seen from Figure 2, the drive test data for the selected performance indicators are sequence data distributed along the railway. In addition, due to the high reliability of the LTE-R network, the constructed dataset for the detection of LTE-R cells with degraded communication performance is a typical imbalanced sequence classification dataset.

3.2. Shapelet and Shapelet Transformation

As defined in [33], shapelets are subsequences of sequence datasets that can maximally distinguish between sequence classes, assuming \( M \) is the length of the shapelet and \( N \) is the length of the sequence in the sequence dataset. Broadly speaking, shapelets are...
not necessarily subsequences of sequence data in the strictest sense. Any sequence that can distinguish between sequence classes in terms of shape and has a length that satisfies $M \leq N$ can be considered as a shapelet of a sequence dataset.

For the convenience of these calculations, the distance between the $i$-th sequence data in a one-dimensional sequence dataset, $X_i$, and the $k$-th shapelet, $S_k$, is defined by Formula (1).

$$d_{i,k} = \min_{j=1,...,J} \frac{1}{M} \sum_{m=1}^{M} (X_{ij+m-1} - S_{k,m})^2, \quad J = N - M + 1$$

In (1), $d_{i,k}$ represents the distance between $S_k$ and $X_i$. Accordingly, the smallest distance between all $J$ subsequences of length $M$ in $X_i$ and $S_k$ is $d_{i,k}$.

After finding the $K$ optimal shapelets in the one-dimensional sequence dataset, the distance between the $K$ optimal shapelets and certain sequence data can be obtained using (1). These distances are used as new features of the sequence data, which map the sequence data to a new feature space. The above process is also called shapelet transformation. Through shapelet transformation, we can transform the one-dimensional sequence dataset $X$ into $D_i$, where $D_i$ is $\{(D_{i1},y_1),(D_{i2},y_2),..., (D_{iN},y_N)\}$, $D_{i1}$ is $\{d_{i1,1},d_{i1,2},...,d_{i1,K}\}$ and the value of $y_i$ is 0 or 1. In general, the value of $K$ is much smaller than $N$, so shapelet transformation can effectively reduce the dimensions of sequence data.

The approach proposed in this paper extracts corresponding shapelets from the drive test data of each LTE-R cell in a supervised manner by modifying the CNN network and performing shapelet transformation to realize feature extraction.

### 4. Methodology

#### 4.1. The Overall Framework of Our Approach

The idea of shapelet transformation is to separate the shapelet extraction process from the classification method. Following dimensionality reduction through shapelet transformation, sequence data can be classified using a general classifier. This also establishes the overall framework of the sequence data classification method, based on shapelet transformation. The proposed feature extraction method for imbalanced sequences follows this framework but also includes the processing of imbalanced sequence data. Therefore, the detection method for LTE-R cells with degraded communication performance proposed in this paper is divided into two steps: (1) Using the modified CNN to extract features from key performance indicator data within the drive test data; (2) Using a general classifier (in this case, SVM) to classify data after feature extraction and achieve the detection of degraded cells. The overall framework of the proposed approach is shown in Figure 3.

From Figure 3, it can be seen that for the one-dimensional sequence dataset composed of the measured values of each key performance indicator from the drive test data, the input one-dimensional sequence data are first passed through a one-dimensional convolutional layer to obtain every subsequence of the input sequence data. Afterward, these subsequences are input into a squared distance layer, in which the weights of the neural network layer represent the shapelets of the input sequence data. Therefore, the role of this layer is equivalent to calculating the distance between each shapelet and each subsequence of the input sequence. After passing through the distance calculation layer, the distance between each shapelet and each subsequence passes through a min pooling layer and the output is the distance between each shapelet and the input sequence data, which is the result of shapelet transformation. As mentioned in Section 3.1, we selected three key performance indicators from the drive test data to use as input data: RSRP, RSRQ, and SINR. In order to perform the shapelet transformation of these three indicator sequences simultaneously, the shapelet transformation results of each indicator sequence are merged in the output layer to form a unified model. In order to obtain shapelets that can maximally distinguish different class sequences, we propose a two-stage training method to train the model. In the first stage, the contrastive loss [34] is used as the loss function to train the feature extraction network. In the second stage, the differential evolution algorithm is introduced to optimize the weights of the squared distance layer while setting an optimization objective based
on the Fisher criterion. The entire model training process can also be regarded as the process of finding the optimal shapelets for each sequence. At this point, feature extraction for the imbalanced multidimensional sequence dataset is complete. Finally, following feature extraction, the dataset is input into an SVM for classification, thereby achieving the detection of LTE-R cells with degraded communication performance.

Figure 3. The overall framework of the proposed approach.

4.2. A Feature Extraction Network-Based CNN

CNNs are important variants of neural networks, which have been widely used in many fields, such as computer vision, natural language processing, and image and sequence feature extraction [35]. CNNs mainly include convolutional layers, pooling layers, activation layers, and other structures. As shown in Figure 3, in this paper, we utilized a one-dimensional convolutional layer and a pooling layer to construct a feature extraction network. The main principles of one-dimensional convolutional layers and pooling layers are as follows.

- **One-dimensional convolutional layer.** Convolutional layers perform discrete convolution operations on input data through convolutional kernels, thereby extracting the features of the input data. Multiple convolution kernels can be used to extract different features of input data. Assuming that \( k = (k_1, k_2, k_3, \ldots, k_M) \) is a one-dimensional convolutional kernel, \( x = (x_1, x_2, x_3, \ldots, x_N) \) is a data record and \( y = (y_1, y_2, y_3, \ldots, y_{N-M+1}) \) is the result of one-dimensional convolutional. The \( j \)-th element of \( y \) is as shown in Formula (2).

\[
y_j = \sum_{m=1}^{M} k_m \cdot x_{j+m-1}
\]

where \( N \) denotes the dimensions of the data record and \( M \) denotes the size of the convolutional kernel.

- **Pooling layers** In CNNs, pooling layers are used to reduce the size of feature maps while expanding the receptive fields of the next-level neural networks. There are several types of pooling layers, including mean pooling, max pooling, stochastic pooling, and min pooling.
The feature extraction network proposed in this article was constructed based on the CNN network structure. In this feature extraction network, the convolutional calculation process of the one-dimensional convolutional network is used to extract all of the specific lengths of the subsequences from the input data. Afterward, we designed a neural network layer that can calculate the distance between each shapelet and subsequence of input data, equal in length to each shapelet. Finally, a min pooling layer is used to obtain the distance between each shapelet and the input data (please refer to Formula (1)).

Assuming that the input sequence length is 500, the number of shapelets is 7, and the shapelet length is 400, the proposed feature extraction network is as shown in Table 2.

Table 2. The structure of the feature extraction neural network.

<table>
<thead>
<tr>
<th>Layer Type</th>
<th>Input Shape</th>
<th>Output Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input layer</td>
<td>(None, 500, 1)</td>
<td>(None, 500, 1)</td>
</tr>
<tr>
<td>1D convolutional layer</td>
<td>(None, 500, 1)</td>
<td>(None, 101, 400)</td>
</tr>
<tr>
<td>Squared distance layer</td>
<td>(None, 101, 400)</td>
<td>(None, 101, 7)</td>
</tr>
<tr>
<td>Min pooling layer</td>
<td>(None, 101, 7)</td>
<td>(None, 7)</td>
</tr>
</tbody>
</table>

The main steps of the feature extraction network proposed in this paper are as follows.

Step 1: Extract the subsequences of the input data. By designing a convolutional kernel for a one-dimensional convolutional neural network, we can extract all subsequences that are the same lengths as the shapelets from the input data. Assuming that the length of the shapelet to be extracted is $M$, the length and number of convolution kernels should also both be $M$. At this point, the $i$-th element of the $j$-th convolution kernel is as shown in Formula (3).

$$k_i^j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

After determining the weights of all convolutional kernels, they are locked, meaning that they do not participate in the subsequent training process. Using the above processing, all subsequences of input data that are of the same length as the shapelets from the input data. For example, when the length of a shapelet is $M$ and the length of the input data sequence is $Q$, we can obtain $Q - M + 1$ subsequences with a length of $M$.

Step 2: Calculate the distance between each shapelet and each subsequence. In order to calculate the distance between each shapelet and the subsequences obtained in Step 1, we designed a squared distance neural network layer to achieve squared distance calculation. The number of weights in this layer depends on the length and number of shapelets. For example, if the length of a shapelet is $M$ and the number of shapelets is $K$, then the number of weights in the squared distance layer is $M \times K$. In fact, these weights can be regarded as the shapelets corresponding to the input data. Therefore, the output of the squared distance layer is the squared distance between each shapelet and each subsequence of input data.

Step 3: Calculate the minimum distance between each shapelet and each subsequence. According to Section 3.2, to obtain the distance between each shapelet and the input data, it is necessary to calculate the minimum distance between each shapelet and each subsequence. The above goals can easily be achieved via the min pooling layer. Figure 4 illustrates the working principles of min pooling.

As can be seen in Figure 4, after passing through the min pooling layer, the output of the squared distance layer can obtain a vector composed of the distance between each shapelet and the input data, i.e., the extracted features.

Through the above steps, the feature extraction network proposed in this paper can complete the shapelet transformation of input data, thereby extracting features.
Figure 4. The working principle of min pooling.

4.3. Optimization Objective for the Feature Extraction of Imbalanced Sequences

Due to the fact that the number of LTE-R cells with degraded communication performance is much smaller than the number of normal cells, the detection of LTE-R cells with degraded communication performance is a typical imbalanced sequence classification problem.

Traditional neural networks for classification often aim to minimize empirical risk during training, such as the cross-entropy loss function. However, when training imbalanced data, the classification information that minority class samples can provide is limited due to the insufficient number of minority class samples. If empirical risk minimization is chosen as the optimization objective of model training, it causes model bias. Therefore, when conducting model training in this paper, we considered optimization objectives based on the Fisher criterion to solve the imbalanced data problem.

The Fisher criterion was proposed by R. A. Fisher in 1936 [36]. The basic idea of the Fisher criterion is to find a certain projection subspace in a feature space so that all feature points can have the best classification performance in that subspace. The basic principle of the Fisher criterion is shown in Figure 5.

Figure 5 shows the projection of a two-dimensional dataset onto a one-dimensional vector. By comparing different projection vectors ($W_1$ and $W_2$), it can be seen that the classification performance of the data samples after projection onto $W_1$ is better than that after projection onto $W_2$. It can also be seen that the basic idea of the Fisher criterion is that the results of a projection function should satisfy the following conditions: the between-class scatter after projection should be as large as possible and the within-class scatter should be as small as possible. In fact, the feature extraction network proposed in this paper can be regarded as a projection function, so the optimization objectives for training the network include the following two aspects:
• Within-class scatter.
Considering that the problem studied in this paper is a binary classification problem and assuming that the two classes are $\Omega_1$ and $\Omega_2$, then $X^l = \{x^l_1, x^l_2, \ldots, x^l_{N_l}\}$ is the number of samples in class $\Omega_l$. So, the within-class scatter matrix of class $\Omega_l$ is as shown in Formula (4).

$$S^l_w = \frac{1}{N_l} \sum_{k=1}^{N_l} \left( x^l_k - m^l \right) \left( x^l_k - m^l \right)^T$$

In (4), $m_l$ is the mean vector of $\Omega_l$. The overall within-class scatter matrix is as shown in Formula (5).

$$S_w = S^1_w + S^2_w$$

The trace of $S_w$ is the overall within-class scatter of $\Omega_1$ and $\Omega_2$. Assuming that $S_w$ is a $M \times M$ matrix, $s_{ij}$ is the element of $S_w$ at row $i$ and column $j$ and $\text{tr}(S_w)$ represents the trace of $S_w$. The trace of $S_w$ is as shown in Formula (6).

$$\text{tr}(S_w) = \sum_{i=1}^{M} s_{ii}$$

• Between-class scatter.
The between-class scatter matrix between $\Omega_1$ and $\Omega_2$ is as shown in Formula (7).

$$S_b = \left( m^1 - m^2 \right) \left( m^1 - m^2 \right)^T$$

The trace of $S_b$ is the between-class scatter between $\Omega_1$ and $\Omega_2$.

After obtaining the within-class scatter and between-class scatter, these two values can be combined by division to form the loss function for training feature extraction networks in order to reach the objectives of minimizing within-class scatter and maximizing between-class scatter. The loss function for training the proposed feature extraction network is as shown in Formula (8).

$$l_{\text{loss}}(\theta) = \frac{\text{tr}(S_w)}{\text{tr}(S_b)}$$

During the model training process, the smaller the value of the loss function, the better the corresponding model performance. By observing the calculation process of this loss function, it can be seen that its calculation results are not affected by the number of samples in each category. Therefore, this loss function can be applied to the classification of imbalanced data.

4.4. Model Training
To make the features extracted by the proposed feature extraction network meet the optimization objectives defined in Section 4.3 as much as possible, we propose a two-stage training method. Figure 6 demonstrates the flow of the two-stage training method.

In the first stage, we utilize the stochastic gradient descent (SGD) optimization method to optimize the model parameters and the loss function is contrastive loss. The SGD optimization method is used to train the feature extraction network proposed in Section 4.2. The contrastive loss is as shown in Formula (9).

$$L \left( W, \left( Y, \bar{X}_1, \bar{X}_2 \right) \right) = \frac{1}{N} \sum_{n=1}^{N} YD^2_W + (1 - Y) \max(m - D_W, 0)^2$$

where $W$ is the weight of the model, $N$ is the size of the training data, $\bar{X}_1$ and $\bar{X}_2$ represent the features of samples $X_1$ and $X_2$, respectively, the value of $Y$ is determined by the categories of samples $X_1$ and $X_2$ (if $X_1$ and $X_2$ do not belong to the same category, then $Y = 0$, ...
otherwise, \( Y = 1 \), \( D w \) is an abbreviation for \( D w (\vec{X}_1, \vec{X}_2) \) (representing the Euclidean distance between \( \vec{X}_1 \) and \( \vec{X}_2 \)), and \( M \) is the set threshold.

**Figure 6.** The flow of the two-stage training method.

As can be seen in (9), when \( Y = 1 \), the model training process can minimize the distance between \( \vec{X}_1 \) and \( \vec{X}_2 \) by adjusting parameters. When \( Y = 0 \) and the distance between \( \vec{X}_1 \) and \( \vec{X}_2 \) is greater than \( m \), there is no need to optimize parameters. When the distance between \( \vec{X}_1 \) and \( \vec{X}_2 \) is less than \( m \), the distance between \( \vec{X}_1 \) and \( \vec{X}_2 \) needs to be increased to \( m \). Therefore, the use of contrastive loss can increase the scatter between different categories.

However, the contrastive loss function is mainly used to increase between-class scatter, without considering within-class scatter. According to the description in Section 4.2, shapelet transformation is mainly completed at the distance calculation layer. Therefore, in the second stage of training, we use the differential evolution (DE) algorithm and the training data to optimize the weights of the distance calculation layer in the first stage of training, with the goals of better meeting the requirements of the Fisher criterion and further improving the classification performance of the model.

The DE algorithm is an optimization algorithm based on swarm intelligence theory, which has the advantages of fast convergence speeds, high reliability, high efficiency, robustness, and strong global search ability [37]. It has been widely used in many optimization problems [38–40]. The main steps of the second stage of training are as follows.

**Step 1: Population initialization.** To achieve better optimization results, individuals in the initialized population should cover the solution space as much as possible. The range of the solution space is determined by the upper \( x_{\text{max}} \) and lower \( x_{\text{min}} \) of weights, while the dimensions of the individual vectors are related to the length and number of shapelets, assuming that the length and number of shapelets are \( M \) and \( K \), respectively. We chose to focus on three key performance indicators from the drive test data, so the dimension of the individual vector is \( D = M \times K \times 3 \) and the \( i \)-th individual in the initial population is \( X_i = \{x_i^1, x_i^2, \ldots, x_i^D\} \), \( i = 1, 2, \ldots, NP \), where \( NP \) is the number of individuals in the initial population. The \( j \)-th element in the individual vector is as shown in Formula (10).

\[
x_j^i = x_{\text{min}} + \text{rand}(0, 1) \cdot (x_{\text{max}} - x_{\text{min}}), \quad j = 1, 2, \ldots, D
\]

**Step 2: Fitness evaluation.** In this paper, in order to calculate the fitness of an individual, firstly, the individual vector needs to be substituted into the distance calculation layer of the feature extraction network as the weights of the distance calculation layer. Then, the training data can be input into the feature extraction network for feature extraction. Finally, the extracted features and corresponding category information are substituted into Formula (8) to obtain the fitness of the individual vector.
Step 3: Mutation. In each round of evolution, a mutation operation is performed on all individuals $X_{i,g}$ in the current generation. If the mutated individual vector is $V_{i,g}$, the mutation operation is as shown in Formula (11).

$$V_{i,g} = X_{a,g} + F \cdot (X_{b,g} - X_{c,g})$$  \hspace{1cm} (11)

In (11), $X_{a,g}$, $X_{b,g}$, and $X_{c,g}$ are three different random individual vectors in the current generation $g$, where $a \neq b \neq c \neq i$, and $F$ is a constant between 0 and 1 that represents the scaling factor.

Step 4: Crossover. Some components of each individual vector $X_{i,g}$ are swapped with the corresponding mutation vector $V_{i,g}$ to obtain the crossover vector $U_{i,g}$. For the $j$-th element $u_{i,g}^j$ of $U_{i,g}$, the crossover operation is as shown in Formula (12).

$$u_{i,g}^j = \begin{cases} v_{i,g}^j \cdot \text{rand}(0,1) < CR \text{ or } j = j_{\text{rand}} \\ x_{i,g}^j \end{cases}$$  \hspace{1cm} (12)

In (12), $CR$ represents the crossover probability (which is a real number between 0 and 1), $D$ is the dimension of the individual vector, and $j_{\text{rand}}$ is a random integer between 0 and 1.

Step 5: Selection. All individual vectors $X_{i,g}$ and their corresponding cross variables $U_{i,g}$ are substituted into the fitness function $l(\cdot)$ to obtain the individual vectors $U_{i,g+1}$ in the next generation. The selection operation is as shown in Formula (13).

$$X_{i,g+1} = \begin{cases} U_{i,g}, l(U_{i,g}) < l(X_{i,g}) \\ X_{i,g}, \text{ else} \end{cases}$$  \hspace{1cm} (13)

Steps 2 to 5 are repeated until the number of iterations reaches the preset maximum value, and then the optimal individual is used as the weight of the distance calculation layer.

Using the above two-stage training method, we obtained a neural network that can extract features from the detection dataset of LTE-R cells with degraded communication performance. Then, simply using traditional classification methods, such as SVM, to classify the data after feature extraction can achieve high detection accuracy.

5. Experiments and Discussion

5.1. Experimental Data

As mentioned in Section 3.1, the experimental data in this paper are drive test data from the Shuo Huang Railway LTE-R network from September 2015 to August 2018, as well as the corresponding EARR data for communication services with a QCI value of 1 exported by the OMC system. Then, we preprocess these data as follows.

Firstly, we extract the data sequences of the key performance indicators (RSRP, RSRQ, and SINR) from the drive test data of each cell, based on the PCI value. Then, using the Lagrangian interpolation to interpolate these three data sequences into sequences with a length of 500, we regard the $3 \times 500$ matrix as a record in the dataset. Finally, we label these records based on the EARR data for communication services with a QCI value of 1 for each cell. If the EARR value is greater than 2%, the cell is marked as an abnormal cell, otherwise, the cell is marked as a normal cell. Through the above process, we obtained the detection dataset of LTE-R cells with degraded communication performance, or the degradation detection dataset for short.

In our experiment, 70% of the data were selected as the training set, and the remaining 30% were used as the test set. Detailed information regarding the degradation detection dataset is shown in Table 3.
Table 3. The detailed information of the degradation detection dataset.

<table>
<thead>
<tr>
<th>Data Properties</th>
<th>Property Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence length</td>
<td>500</td>
</tr>
<tr>
<td>Dataset size</td>
<td>2554</td>
</tr>
<tr>
<td>Normal data size</td>
<td>2470</td>
</tr>
<tr>
<td>Abnormal data size</td>
<td>84</td>
</tr>
<tr>
<td>Training set size</td>
<td>1788</td>
</tr>
<tr>
<td>Test set size</td>
<td>766</td>
</tr>
<tr>
<td>Imbalance ratio</td>
<td>29.4</td>
</tr>
</tbody>
</table>

5.2. Experiment Settings

Firstly, we constrain the weights to be optimized in the feature extraction network so that they fall in the range of $-1$ to $1$, which also limits the search range of the differential evolution algorithm. Then, in the first stage of model training, the learning rate of the SGD optimization method is set as 0.01. Furthermore, in the feature extraction network, there are two parameters that can directly affect the effectiveness of feature extraction: the length and number of shapelets, assuming that the length of shapelets is $M$ and the number of shapelets is $K$. To find the appropriate parameter values, we conducted a grid search on $M \in \{50, 100, 150, 200, 250, 300, 350, 400, 450, 500\}$ and $K \in \{3, 4, 5, 6, 7, 8, 9, 10\}$. Due to the fact that the F1 score can comprehensively consider precision and recall, which are important evaluation metrics in detecting LTE-R cells with degraded communication performance [3], the F1 score is used as the evaluation metric for the grid search. When conducting the grid search, an SVM is used to classify the training set after feature extraction. The average F1 score obtained from a 5-fold cross-validation is taken as the evaluation result for each search. The grid search results for $M$ and $K$ are shown in Figure 7.

![Figure 7](image_url)

Figure 7. The grid search results of $M$ and $K$. (a) shows the grid search results for $K$ and $M$ in the form of a heat map, and the darker the color, the higher the F1 score. (b) shows the grid search results for $K$ and $M$ in the form of a group bar chart, and the longer the length of the bar, the higher the F1 score.

It can be seen from Figure 7 that when $M$ equals 300 and $K$ equals 5, we can obtain the maximum F1 score value. Therefore, in this experiment, we set the length of the shapelets $M$ to 300 and the number of shapelets to 5.

After determining the length and number of shapelets, it is necessary to determine the relevant parameters of the differential evolution algorithm in order to further ensure the performance of the detection method, i.e., the maximum iteration $G$ and the population $NP$. Generally speaking, the larger the population size and maximum generation, the better the optimization effect. However, increasing $G$ and $NP$ can also lead to increased computational complexity, which affects algorithm efficiency. To find appropriate parameter values, we conduct a grid search on $G \in \{0, 40, 60, 80, 100\}$ and $NP \in \{10, 20, 30, 40, 50\}$. Similarly, we conduct a grid search and utilize an SVM to classify the training set after feature extraction. The average F1 score obtained from a 5-fold cross-validation is taken as
the evaluation result for each search. The grid search results for $G$ and $NP$ are shown in Figure 8.

![Figure 8](image)

**Figure 8.** The grid search results of $G$ and $NP$. (a) shows the grid search results for $G$ and $NP$ in the form of a heat map, and the darker the color, the higher the F1 score. (b) shows the grid search results for $G$ and $NP$ in the form of a group bar chart, and the longer the length of the bar, the higher the F1 score.

As can be seen from Figure 8, when $NP$ equals 30 and $G$ equals 60, we can obtain the maximum F1 score. Therefore, in this experiment, we set the population to 30 and the maximum iteration to 60.

All the codes of our approach and comparison methods are written in Python 3.6. All the experiments were carried out on a laptop, which was configured as follows: CPU is i5-8400 (Intel, Santa Clara, CA, USA), memory is 8G, and the graphics card is NVIDIA 1050Ti 4G (NVIDIA, Santa Clara, CA, USA).

### 5.3. Comparison and Discussion

In this section, comparative experiments on the detection dataset of LTE-R cells with degraded communication performance are presented in order to verify the performance of the method proposed in this paper. The experimental results are discussed and analyzed.

Before conducting the comparative experiments, we first verified the effectiveness of the proposed two-stage training method on the experimental dataset. Figure 9 illustrates the classification performance of the model after training it 10 times on the training and test sets, both with and without using the two-stage training method.

![Figure 9](image)

**Figure 9.** Comparison of different training methods.

As can be seen from Figure 9, the two-stage training method proposed in this paper can significantly improve the training performance of the model compared to the training method that only uses contrastive loss and SGD (one-stage training). This is because the
use of the differential evolution algorithm compensates for the deficiency of contrastive loss, which only focuses on between-class scatter.

When conducting the comparative experiments, several imbalanced data classification methods, such as SMOTE-SVM [41], ADASYM-SVM [41], Cosen-SVM [42], and several anomaly detection methods, such as LOF [7], VAE [14], and DEDAE [3] are selected as comparison methods. Considering that the detection of LTE-R cells with degraded communication performance is a sequence data classification problem, we also selected two sequence data classification approaches for performance comparison: the fast shapelet tree (FST) proposed by Rakthanmanon et al. [43] and the learning time series (LTS) proposed by Grabocka et al. [44]. Furthermore, considering the imbalance of the experimental data, a CNN-based imbalanced sequences classification approach, called the cost-sensitive convolution neural network (CSCNN) [26], was also selected as a comparison approach.

To verify the effectiveness of the proposed approach for feature extraction, we designed experiments to compare the feature extraction capabilities. Figure 10 shows the distribution of features extracted by our approach, FST, LTS, CSCNN, VAE, and DEDAE on the degradation detection dataset in the feature space.

Figure 10 illustrates the feature data after PCA dimensionality reduction. It can be seen that the features extracted by the proposed feature extraction neural network were able to distinguish between normal and abnormal cells at the feature level. This is because the proposed optimization objective, based on the Fisher criterion, can minimize the within-class scatter of extracted features while maximizing the between-class scatter. The features extracted by the other five comparison approaches could not clearly distinguish between the two types of samples at the feature level, which affected the accuracy of classification. The experiments showed that the proposed approach has obvious advantages over the comparison approaches when extracting imbalanced sequence features.

Then, in order to verify the classification performance of the proposed approach, we compared the classification performance of our model to baselines on the degradation detection dataset. The evaluation metrics for classification performance include G-mean, AUC, F1 score, and precision. The calculation process of each evaluation metric can be found in [45]. During the experiments, each approach trained the model on the training set and used the trained model to calculate the various evaluation metrics on the test set. After repeating the above process five times, we could obtain the mean and standard deviation of the different evaluation metrics for each approach. Table 4 shows the classification performance of each approach on the degradation detection dataset.
Table 4. The classification results of each approach.

<table>
<thead>
<tr>
<th>Name</th>
<th>G-Mean ± Standard Deviation</th>
<th>AUC ± Standard Deviation</th>
<th>F1 Score ± Standard Deviation</th>
<th>Precision ± Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMO-T-SVM</td>
<td>0.7194 ± 0.1048</td>
<td>0.7238 ± 0.1079</td>
<td>0.7041 ± 0.1194</td>
<td>0.6862 ± 0.1028</td>
</tr>
<tr>
<td>ADASYM-SVM</td>
<td>0.7063 ± 0.1035</td>
<td>0.7095 ± 0.1066</td>
<td>0.6919 ± 0.1195</td>
<td>0.6595 ± 0.0857</td>
</tr>
<tr>
<td>Cosen-SVM</td>
<td>0.6835 ± 0.1360</td>
<td>0.6929 ± 0.1339</td>
<td>0.6658 ± 0.1637</td>
<td>0.6483 ± 0.1385</td>
</tr>
<tr>
<td>FST</td>
<td>0.5231 ± 0.0058</td>
<td>0.5333 ± 0.0105</td>
<td>0.4788 ± 0.0009</td>
<td>0.501 ± 0.0026</td>
</tr>
<tr>
<td>LTS</td>
<td>0.5294 ± 0.0001</td>
<td>0.5292 ± 0.0001</td>
<td>0.5262 ± 0.0087</td>
<td>0.5195 ± 0.0007</td>
</tr>
<tr>
<td>CSCNN</td>
<td>0.8448 ± 0.0332</td>
<td>0.8564 ± 0.0277</td>
<td>0.7439 ± 0.0263</td>
<td>0.7741 ± 0.0309</td>
</tr>
<tr>
<td>LOF</td>
<td>0.6217 ± 0.1667</td>
<td>0.6405 ± 0.158</td>
<td>0.5911 ± 0.19</td>
<td>0.6467 ± 0.204</td>
</tr>
<tr>
<td>VAE</td>
<td>0.7101 ± 0.1215</td>
<td>0.719 ± 0.1093</td>
<td>0.6793 ± 0.1499</td>
<td>0.7029 ± 0.1175</td>
</tr>
<tr>
<td>DEDAE</td>
<td>0.8032 ± 0.1034</td>
<td>0.8126 ± 0.0428</td>
<td>0.6824 ± 0.0637</td>
<td>0.7994 ± 0.1132</td>
</tr>
<tr>
<td>Our approach</td>
<td>0.9863 ± 0.0092</td>
<td>0.9864 ± 0.0091</td>
<td>0.9799 ± 0.0004</td>
<td>0.9872 ± 0.0181</td>
</tr>
</tbody>
</table>

In order to compare the classification performance of the comparison approaches on the degradation detection dataset more intuitively, Figure 11 shows the performance of each approach on different evaluation metrics.

![Performance of each approach on different evaluation metrics.](image)

As can be seen from Figures 10 and 11, our approach was significantly superior to all other comparison approaches in terms of G-mean, AUC, F1 score, and precision. CSCNN also achieved good results, which can be attributed to its use of cost-sensitive loss functions and the strong feature extraction ability of convolutional networks for sequence data. However, CSCNN still had a significant disadvantage compared to our approach due to the difficulty in determining cost-sensitive factors. The imbalanced data classification methods and anomaly detection methods also showed relatively good performance. However, as can be seen from Figure 11, these methods were significantly less stable than our approach. FST and LTS performed the worst due to the insufficient information provided by the minority class samples, resulting in model bias.

From the above experiments, it can be seen that the imbalanced sequence feature extraction neural network proposed in this paper has excellent performance for the detection of LTE-R cells with degraded communication performance. Our approach can achieve an accuracy of about 98% on various imbalanced data classification evaluation metrics, which provides good decision support for the operation and maintenance of the Shuo Huang Railway LTE-R communication network.
6. Conclusions

In order to solve the problem of detecting LTE-R cells with degraded communication performance in the Shuo Huang Railway, we designed a deep neural network structure that can achieve imbalanced sequence feature extraction based on the CNN network structure and the idea of shapelet transformation. Then, considering the imbalance between the numbers of degraded cell samples and normal cell samples, the optimization objective of the imbalanced sequence feature extraction neural network was designed based on the Fisher criterion. Afterward, in order to determine the appropriate weights of the feature extraction network, we developed a two-stage training method. In the first stage, we used SGD and contrastive loss to preliminarily train the model. In the second stage, we used the differential evolution algorithm to further optimize the weights of the distance calculation layer. Finally, based on the sequences of key performance indicators from the drive test data for each Shuo Huang Railway LTE-R cell, as well as the corresponding EARR data, we constructed a degradation detection dataset and conducted experiments on this dataset using the proposed approach and other comparison approaches. The experimental results show that the proposed feature extraction network could effectively extract the features of imbalanced sequences and that its classification performance was significantly better than that of the other comparison approaches. The proposed approach achieved an accuracy of 98% or above in detecting LTE-R cells with degraded communication performance, which means that it could be a powerful tool for the operation and maintenance of the Shuo Huang Railway LTE-R network and has high application value.

Author Contributions: Conceptualization, J.Q. and H.M.; methodology, J.Q.; software, J.Q.; investigation, J.Q. and H.M.; resources, J.Q.; data curation, J.Q.; writing—original draft preparation, J.Q.; writing—review and editing, J.Q., H.M. and C.Q.; visualization, J.Q.; supervision, C.Q.; project administration, J.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Key Scientific and Technological Project of Henan Province (no. 222102210006) and the Science and Technology Development Project of China Railway Design Corporation (no. 2022A02538005).

Data Availability Statement: The data can be shared upon request and the data are not publicly available due to information security issues.

Conflicts of Interest: Jiantao Qu, Chunyu Qi and He Meng are employees of China Railway Design Corporation (China) Company. The authors declare no conflicts of interest regarding the publication of this research article.

References


42. Iranmehr, A.; Masnadi-Shirazi, H.; Vasconcelos, N. Cost-sensitive support vector machines. *Neurocomputing* **2019**, *343*, 50–64. [CrossRef]


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.