Application of the SGT Model in Magnetic Signal Anomaly Detection and Improvement Approaches

Authors: Hu Suhang

Date: 2024-01-06T00:00:00+00:00

Abstract

This report investigates the application of the SGT model in the field of magnetic detection, with particular attention to its performance on the MGT, SNR0, and SNR5 datasets. Experimental results reveal that the SGT model suffers from issues such as excessively high false alarm rates and large prediction biases when processing these datasets. To address the insufficient predictive capability and generalization ability of the model, we designed a series of improvement experiments, focusing on three aspects: parameter tuning, optimizing feature extraction methods, and modifying continuity determination. Among these three improvement methods, parameter tuning achieved approximately a 0.5% performance improvement, while the feature extraction optimization and orthogonal basis judgment methods actually decreased prediction performance by 20%. Through code review and logical reasoning, we identified that the problem stems from incompatibility between feature extraction and the model. To adapt to the orthogonal basis algorithm, we propose an improved approach: introducing multiple different types of features, including time-domain features, frequency-domain features, and statistical features, among others, and comprehensively utilizing this feature information to construct a more complex and comprehensive SGT model. Furthermore, a stacking module is introduced, which takes the prediction results of single models based on different features as input and generates more accurate predictions through further learning and integration.

Full Text

Preamble

Beijing University of Chemical Technology

Title: Application and Improvement Strategies of the SGT Model in Magnetic Signal Anomaly Detection

Author: Hu Suhang

Major: Computer Science and Technology

College: College of Information Science and Technology

Date: January 6, 2024

Application and Improvement Strategies of the SGT Model in Magnetic Signal Anomaly Detection

This report investigates the application of the SGT model in the field of magnetic prospecting, with particular attention to its performance on the MGT, SNR0, and SNR5 datasets. Experimental results reveal that the SGT model suffers from high false alarm rates and significant prediction bias when processing these datasets. To address the model's insufficient predictive and generalization capabilities, we designed a series of improvement experiments focusing on three aspects: parameter tuning, optimizing feature extraction methods, and modifying continuity judgment.

Among these three improvement approaches, parameter tuning achieved approximately 0.5% performance improvement, while feature extraction optimization and orthogonal basis judgment methods 反而 reduced prediction effectiveness by 20%. Through code review and logical reasoning, we identified that the problem stems from incompatibility between feature extraction and the model. To adapt to the orthogonal basis algorithm, we propose an improvement strategy: introducing multiple different types of features, including time-domain features, frequency-domain features, and statistical features, and comprehensively utilizing this feature information to construct a more complex and comprehensive SGT model. Additionally, we introduce a stacking module that takes the prediction results of single models based on different features as input and generates more accurate predictions through further learning and synthesis.

Keywords: SGT, Magnetic Prospecting, Multi-feature Prediction

Abstract

This report explores the application of the SGT model in the field of magnetic prospecting, with special focus on its performance on the MGT, SNR0, and SNR5 datasets. The experimental results reveal that the SGT model suffers from high false alarm rates and large prediction bias when dealing with these datasets. To address the insufficient predictive and generalization abilities of the model, we designed a series of improvement experiments focusing on three aspects: parameter tuning, optimizing the feature extraction method, and modifying the continuity judgment.

Among these three improvement methods, parameter tuning achieved about 0.5% performance improvement, while the methods of feature extraction optimization and orthogonal basis judgment instead reduced the prediction effect by 20%. Through code review and logical reasoning, we found that the problem stems from feature extraction incompatibility with the model. In order to adapt to the orthogonal basis algorithm, we propose an improvement idea: introduce many different types of features, including time-domain features, frequency-domain features, and statistical features, etc., and comprehensively utilize the information of these features to construct a more complex and comprehensive SGT model. In addition, the stacking module is introduced to take the prediction results of a single model based on different features as inputs, and generate a more accurate ultimate prediction through further learning and synthesis.

KEY WORDS: SGT, Magnetic Prospecting, Multi-feature Prediction

---

Table of Contents

---

1. Introduction

Magnetic anomaly detection plays a key role in geoscience and resource exploration. The difficulty of magnetic anomaly detection technology lies in the fact that magnetic anomaly signals are typically very weak and susceptible to interference from geomagnetic background noise, instrument noise, and electromagnetic interference. Therefore, effective feature extraction, classification, and detection of magnetic anomaly signals are essential. The SGT model improves upon the Vision Transformer architecture by modifying the patch split construction, enabling the combination of Transformer with the Shifted-Grad Block, which enhances long-range dependency modeling for one-dimensional data and demonstrates promising performance on magnetic data. This paper aims to thoroughly investigate the specific application of the SGT model[1] in magnetic signal detection and briefly discuss some challenges and problems identified in practical applications.

---

2. Magnetic Signal Anomalies

2.1 Basic Concepts of Magnetic Signal Anomalies

Magnetic signal anomalies typically refer to phenomena where the Earth's magnetic field exhibits abnormal variations in a certain region or time period. The Earth's magnetic field is generated by the flow of liquid iron in the planet's outer core, making it dynamic and complex. Magnetic signal anomalies may be caused by subsurface geological structure changes, mineral resource distributions, crustal movements, and other factors[2][3]. Monitoring and analyzing magnetic signal anomalies hold significant value in geoscience and exploration fields. Magnetic anomaly detection helps identify subsurface geological features such as faults, ore deposits, and rock layer variations, with practical applications in mineral exploration, geological surveys, and natural disaster early warning. Detecting magnetic signal anomalies typically requires high-precision magnetometers or magnetic instruments to measure and analyze the geomagnetic field for identifying anomalies. In magnetic signal anomaly detection research, various mathematical models, machine learning methods, or deep learning techniques are often employed to improve the accuracy and sensitivity of anomaly signal detection, thereby enhancing understanding of subsurface structures and geodynamic processes.

2.2 Principles of Magnetic Signal Anomaly Detection

The Earth's geomagnetic field, detectable at the surface, is generated by free electron currents and ferromagnetic materials in the planet's core. The surface geomagnetic field can typically be equivalently modeled as a magnetic dipole field. Magnetic anomaly detection technology is based on the principle that ferromagnetic objects entering the original electromagnetic field alter its distribution. For instance, ferromagnetic targets such as vehicles, ships, and aircraft change the original electromagnetic distribution of the geomagnetic field during motion, creating what is known as a magnetic anomaly. Figure 2-[Figure 2: see original paper]2 illustrates the magnetic anomaly disturbance caused by a ferromagnetic target.

Since the geometric shapes of magnetic targets are complex, direct analysis is extremely difficult and lacks generalizability. Theoretical studies indicate that when the detection spatial length exceeds 2.5 times the size of the magnetic anomaly target, it can be equivalently analyzed as a magnetic dipole model. In practical experimental scenarios and real-world environments, detectors typically fly at altitudes of several hundred meters, far exceeding 2.5 times the magnetic anomaly target size.

Using the ferromagnetic target as the origin and the aircraft flight direction as the x-axis, we establish a model as shown in Figure 2-[Figure 2: see original paper]3, where M represents the target's magnetic moment, T denotes the local geomagnetic field vector, r represents the distance vector from the target to the detector, d indicates the shortest distance from the target to the detector's trajectory, and I and D represent the inclination and declination angles of the target's magnetic moment, respectively.

The form of the magnetic anomaly signal generated by the target is:

[Equation content would appear here based on Equation (2-1)]

In practical applications, scalar magnetometers are typically used to detect scalar magnetic fields for magnetic anomaly detection. In theoretical research, the vector B is the sum of the magnetic dipole field B_{dipole} and the geomagnetic field T:

B = B_{dipole} + T ……………………………………………. Equation (2-2)

The corresponding scalar magnetic field is:

|B| = |B_{dipole} + T| ……………………………………… Equation (2-3)

Since the geomagnetic field intensity is much greater than the magnetic dipole field, Equation (2-3) can be approximated as:

|B| |T| + (B_{dipole} · T)/|T| ……………………………… Equation (2-4)

When |B_{dipole}| |T|, based on the approximation relationship, we can further approximate the above equation (|B_{dipole}|/|T| 1) as the scalar magnetic field of the geomagnetic field. Therefore, the scalar magnetic field generated by the magnetic dipole is the projection of the vector field in the direction of the local geomagnetic field:

ΔB B_{dipole} · T/|T| ……………………………………………………. Equation (2-5)

This can be expressed as:

ΔB (_0/4π|r|^3)[3(M · r)r/|r|^2 - M] · T/|T| ………………………………….. Equation (2-6)

2.3 Magnetic Signal Anomaly Detection Technology

Magnetic signal anomaly detection technology is a non-contact sensing method that uses magnetometers or magnetic gradiometers to measure changes in the geomagnetic field, thereby identifying and locating magnetic targets or geological structures. This technology has broad applications in geological exploration, military reconnaissance, archaeological excavation, environmental monitoring, and other fields. The challenge of magnetic signal anomaly detection technology is that magnetic anomaly signals are typically very weak and susceptible to interference from geomagnetic background noise, instrument noise, and electromagnetic interference. Therefore, effective feature extraction, classification, and detection of magnetic anomaly signals are necessary.

Traditional magnetic signal anomaly detection technologies are mainly divided into two categories: time-domain methods and frequency-domain methods. Time-domain methods directly analyze the time-domain waveform of magnetic anomaly signals, such as orthogonal basis function decomposition and minimum entropy detectors. Orthogonal basis function decomposition represents the magnetic anomaly signal as a linear combination of a set of orthogonal basis functions, and then determines the presence of magnetic anomalies based on the basis function coefficients. The minimum entropy detector uses the entropy value of the magnetic anomaly signal as a detection statistic; when the entropy value falls below a certain threshold, a magnetic anomaly is considered present. These methods offer theoretical interpretability and guarantees but suffer from high false alarm rates and difficulties in modeling geomagnetic noise[4].

Frequency-domain methods first apply Fourier transform to magnetic anomaly signals to convert them into frequency-domain signals, and then analyze the characteristics of the frequency-domain signals, such as higher-order zero-crossing detectors. Higher-order zero-crossing detectors use the zero-crossing points of the higher-order spectrum of magnetic anomaly signals as detection statistics; when the number of zero-crossing points exceeds a certain threshold, a magnetic anomaly is considered present[5]. These methods offer strong noise resistance and high detection sensitivity but require complex data preprocessing and spectrogram generation, affecting inference speed.

Although traditional magnetic signal anomaly detection technologies have achieved magnetic anomaly signal detection to some extent, they still have limitations, such as insufficient feature extraction, inaccurate classification and detection, and imprecise parameter estimation. To overcome these limitations, recent years have seen the emergence of machine learning-based magnetic signal anomaly detection technologies, such as support vector machines, convolutional neural networks, and recurrent neural networks. These techniques utilize deep learning[6] for feature extraction and classification of magnetic anomaly signals, offering generalization capabilities and adaptability, but also requiring complex data preprocessing and spectrogram generation, which affects inference speed.

In summary, traditional magnetic signal anomaly detection technology is a fast and effective non-contact sensing method, but it has defects and limitations that require further research and improvement to enhance the accuracy and speed of magnetic signal anomaly detection[7][8][9][10].

3. SGT Model

3.1 Introduction to the SGT Model

The SGT model is a one-dimensional continuous time series data analysis method based on Transformer, consisting primarily of two components: the Shifted-Grad block and an improved Transformer encoder. The Shifted-Grad block is a module for extracting gradient features from one-dimensional data, comprising a shift module, difference module, raw data feature fusion module, gradient embedding module, multi-scale module, and feature smoothing module. These modules construct gradient features from nearest neighbors to shifted nearest neighbors by controlling data shift times, nearest-neighbor relationships, and distance dependencies, simplifying data preprocessing steps and enhancing data representation capabilities. The improved Transformer encoder is a module for capturing long-range dependencies in one-dimensional data. Based on the Vision Transformer model, it improves the patch partitioning method, enabling combination with the Shifted-Grad block and enhancing classification and detection performance for one-dimensional data.

The SGT model has been experimentally validated on multiple one-dimensional continuous time series data tasks, including magnetic anomaly signal classification and detection, remote sensing image classification, and arrhythmia diagnosis. Compared with existing methods, the SGT model not only achieves higher accuracy and recall rates but also offers faster inference speeds and fewer model parameters. The SGT model provides a general framework for processing one-dimensional continuous time series data and represents an innovative approach. Its structure is shown in Figure 3-[Figure 3: see original paper]1.

3.2 Mathematical Principles of the SGT Model

The SGT model is an analysis method for one-dimensional continuous time series data based on Transformer, consisting primarily of two parts: the SG block and an improved Transformer. The SG block is a module for extracting gradient features from one-dimensional data, comprising a shift module, difference module, raw data feature fusion module, gradient embedding module, feature smoothing module, and multi-scale module. The improved Transformer is a module for enhancing long-range dependencies in one-dimensional data, modifying the patch partitioning method based on Vision Transformer to preserve sequence information.

The mathematical principles of the SG block are based on the concept of mathematical differences, which are used to approximate function derivatives and detect change rates and mutation points in data. The SG block takes a one-dimensional data vector x as input and outputs a gradient feature matrix G. The computation process of the SG block is as follows:

a) **Shift Module**: The shift module controls the number and direction of data shifts to construct different neighborhood relationships and distance dependencies. Its input is x, and its output is a shift matrix S of size n×s, where n is the number of shifts and s is the data length. The shift module's calculation formula is:

$$S_{i,j} = x_{j-i}$$

where i represents the shift count and j represents the data index. Each row of S is a shifted version of x.

b) **Difference Module**: The difference module calculates first-order and second-order differences of the data to extract change rates and curvature. Its input is S, and its output is a difference matrix D of size 2n×(s-n-1). The difference module's calculation formula is:

$$D_{i,j} = S_{i,j+1} - S_{i,j}$$

$$D_{i+n,j} = S_{i,j+2} - 2S_{i,j+1} + S_{i,j}$$

where i represents the shift count and j represents the data index. The first n rows of D are the first-order differences of S, and the last n rows are the second-order differences of S.

c) **Raw Data Feature Fusion Module**: This module retains raw data information to avoid losing certain effective information. Its inputs are x and D, and its output is a fusion matrix R of size (2n+1)×(s-n-1). The calculation formula is:

$$R_{i,j} = \begin{cases} x_{j+n/2} & \text{if } i = 0 \\ D_{i,j} & \text{if } 1 \leq i \leq 2n \end{cases}$$

where i represents the feature dimension and j represents the data index. The first row of R is a truncated version of x, and the remaining 2n rows are copies of D.

d) **Gradient Embedding Module**: This module integrates multi-order difference relationships to enhance gradient features. Its input is R, and its output is a gradient feature matrix G of size n×(s-n-1). The calculation formula is:

$$G_{i,j} = \sum_{k=0}^{2n} W_{i,k} \cdot R_{k,j}$$

where i represents the feature dimension, j represents the data index, and W is a learnable weight matrix of size n×(2n+1) used to perform weighted summation on each row of R to obtain each row of G.

e) **Feature Smoothing Module**: This module makes features smoother to eliminate noise and outliers. Its input is G, and its output is a smoothed feature matrix H of size n×(s-n-1). The calculation formula is:

$$H_{i,j} = \frac{1}{2m+1} \sum_{k=-m}^{m} G_{i,j+k}$$

where i represents the feature dimension, j represents the data index, and m is a hyperparameter representing the radius of the smoothing window. Each element of H is a local mean of G.

f) **Multi-scale Module**: This module extracts features at different scales to enhance diversity and robustness. Its input is H, and its output is a multi-scale feature matrix M of size n×(s-n-1). The multi-scale module's calculation formula is:

$$M_{i,j} = \sum_{k=1}^{n} V_{i,k} \cdot H_{k,j}$$

where i represents the feature dimension, j represents the data index, and V is a learnable weight matrix of size n×n used to perform weighted summation on each row of H to obtain each row of M.

3.3 Advantages and Limitations of the SGT Model

The SGT model is a one-dimensional continuous time series data analysis method based on Transformer with the following advantages and limitations:

Advantages: 1. **Simplified Data Preprocessing**: The SGT model extends the gradient feature space of one-dimensional data through shift and difference modules, eliminating tedious preprocessing operations such as normalization, filtering, and spectrogram conversion. 2. **Improved Classification Accuracy and Speed**: Through raw data feature fusion, gradient embedding, multi-scale modules, and feature smoothing, the SGT model enhances feature representation capabilities for one-dimensional data. By improving the patch partitioning method of Transformer, it preserves sequence integrity and long-range dependencies. On multiple one-dimensional continuous datasets, the SGT model achieves higher classification accuracy than existing methods while offering faster inference speeds and fewer parameters. 3. **Enables Target Detection**: The SGT model transforms the target detection task for one-dimensional continuous data into a classification task through a sliding window approach, then eliminates false alarms in sliding window detection results through post-processing dynamic programming algorithms, achieving efficient target detection.

Limitations: 1. **Not Suitable for Discontinuous Data**: The SGT model's fundamental assumption is that one-dimensional data has continuity, meaning changes between data points are smooth. Therefore, for some discontinuous data such as speech signals, the SGT model's performance decreases, requiring data transformation or model modifications. 2. **High Parameter Tuning Requirements**: The SGT model's performance is affected by parameter settings in the shift and difference modules. These parameters need adjustment for different datasets; otherwise, they may lead to insufficient or excessive feature extraction.

3.4 Effectiveness of the SGT Model in Magnetic Signal Anomaly Detection

Classification Accuracy: The SGT model achieved 99.01% accuracy on magnetic signal classification tasks, surpassing the best baseline method by 1.57 percentage points, demonstrating its effectiveness in extracting magnetic signal features and performing classification.

Inference Speed: The SGT model can process 2.8$×10^{3}$ inputs per second, nearly 28 times faster than the best baseline method, indicating high computational efficiency and real-time capability.

Detection Effectiveness: Through sliding windows and post-processing dynamic programming algorithms, the SGT model achieves target detection for magnetic signals, accurately locating the start and end points of magnetic anomaly signals without false alarms or missed detections, demonstrating its effectiveness in handling magnetic signal detection tasks.

3.5 Specific Application Method of the SGT Model in Magnetic Signal Anomaly Detection

The SGT model is applied to magnetic signal anomaly detection through the following process:

First, magnetic signals are segmented into sequences of 501 points as input to the SGT model. This step transforms one-dimensional magnetic signals into a form suitable for SGT processing—an n×s matrix where n is the feature dimension and s is the sequence length. A length of 501 was chosen because it can cover most magnetic anomaly signals without introducing excessive redundant information.

Next, the SGT model classifies each sequence to determine whether it contains magnetic anomaly signals. This step leverages the SGT model's classification capability to perform binary classification on each sequence. The SGT model's output is an n×(s-n-1) matrix, which passes through a fully connected layer and a softmax layer to produce a 2×(s-n-1) matrix representing classification probabilities at each position. A classification threshold of 0.5 was adopted, where positions with magnetic anomaly signal probability exceeding 0.5 are considered anomalous.

When magnetic anomaly signals are detected, the sequence is further segmented into sub-sequences of 51 points using a sliding window as input to the SGT model. This refinement step helps determine the location and extent of magnetic anomaly signals. A sub-sequence length of 51 ensures each sub-sequence contains at least one magnetic anomaly signal while avoiding excessive overlap. The sliding window approach with overlapping regions prevents magnetic anomaly signals from being split between windows.

The SGT model then classifies each sub-sequence to determine magnetic anomaly presence, using the same 0.5 threshold. Finally, a post-processing dynamic programming algorithm eliminates false alarms and determines the precise location and range of magnetic anomaly signals. This algorithm treats classification probabilities at each position as state values and magnetic anomaly presence as state transition conditions to construct a state transition graph. Dynamic programming finds the optimal path maximizing the sum of state values to obtain final magnetic anomaly positions and ranges. A hyperparameter specifying the minimum magnetic anomaly signal length constrains the state transition graph to improve algorithm efficiency.

4. Experimental Design and Results Analysis

4.1 Experimental Design

Experimental Objective: To explore the performance of the SGT model on MGT, SNR0, and SNR5 datasets, identify existing problems, and determine improvement directions.

Dataset Preparation:
- **MGT Dataset**: Contains magnetic signal data from different regions, times, and depths, with 40,000 samples, each having 12 features.
- **SNR0 Dataset**: Based on the MGT dataset with added Gaussian noise at 0 dB signal-to-noise ratio, containing 40,000 samples with 12 features each.
- **SNR5 Dataset**: Based on the MGT dataset with added Gaussian noise at 5 dB signal-to-noise ratio, containing 40,000 samples with 12 features each.

Experimental Steps:
1. Data preprocessing: Perform necessary preprocessing on MGT, SNR0, and SNR5 datasets, including data cleaning and normalization.
2. Model training: Train the SGT model separately on each dataset, recording the training process including loss function values and accuracy per epoch.
3. Model validation: Validate the model using a validation set and record results.
4. Model testing: Test the model using a test set and record results.
5. Performance evaluation: Assess model performance using metrics such as false alarm rate, miss rate, and accuracy.
6. Results analysis: Conduct in-depth analysis of experimental results, including model performance across datasets, defects and errors, potential causes, and solutions.

4.2 Experimental Implementation

Experimental Environment:
- Deep Learning Framework: PyTorch
- Operating System: Windows
- GPU: NVIDIA GeForce RTX 4070ti
- Programming Language: Python

Dataset Preparation: Download MGT, SNR0, and SNR5 datasets from the magnetic prospecting dataset and save them as mgt.csv, snr0.csv, and snr5.csv files. Each file contains 40,000 rows with 12 comma-separated values per row, representing features of a magnetic signal sample.

Data Preprocessing: For each dataset, perform the following operations:
1. Read the CSV file and convert data to numpy arrays with shape (40000, 12).
2. Normalize each sample's features to have zero mean and unit standard deviation.
3. Split datasets into training, validation, and test sets in an 8:1:1 ratio, randomly shuffling to ensure uniform distribution.

4. Save datasets as torch.tensor objects for convenient model training and testing.

Model Training: Train the SGT model on each dataset as follows:
1. Define the SGT model structure and parameters, using Vision Transformer as the base and improving the patch split construction to combine Transformer with Shifted-Grad Block, enhancing long-range dependencies for one-dimensional data.
2. Define loss function and optimizer, using cross-entropy loss and Adam optimizer with learning rate 0.001, weight decay 0.0001, and batch size 64.
3. Train the model for a maximum of 200 epochs. After each epoch, evaluate accuracy on the validation set. Training stops if accuracy does not improve for 5 consecutive epochs or reaches the maximum, saving the best model parameters.
4. Record the training process, including loss and accuracy per epoch, plot training curves, and analyze convergence and generalization.

Model Validation: Validate the model using the validation set:
1. Load the best model parameters, set the model to evaluation mode, and disable gradient computation.
2. Perform forward propagation on validation data, calculate accuracy, and record validation results.
3. Analyze model performance on the validation set to identify strengths and weaknesses for reference in subsequent testing and analysis.

Model Testing: Test the model using the test set:
1. Load the best model parameters, set the model to evaluation mode, and disable gradient computation.
2. Perform forward propagation on test data, calculate accuracy, and record test results.
3. Analyze model performance on the test set, compare with validation results, and evaluate stability and robustness.

Performance Evaluation: Assess model performance using accuracy, false alarm rate, and miss rate:
1. For each dataset, compute confusion matrices on the test set and calculate accuracy, false alarm rate, and miss rate.
2. Compare performance across datasets to analyze detection capability for magnetic signals at different SNR levels and evaluate model strengths and weaknesses.

Results Analysis: Conduct in-depth analysis including model performance across datasets, defects and errors, potential causes, and solutions:
1. Visualize model performance across datasets using line charts and scatter plots to show training processes, validation results, test results, and performance metrics, analyzing trends and characteristics.
2. Locate model defects and errors by showing error distributions across different categories, SNR levels, and features to analyze error sources and influencing

factors.
3. Investigate potential causes using correlation analysis, principal component analysis, and clustering to show relationships between dataset features and labels, analyzing dataset difficulty and complexity as well as model adaptability and limitations.
4. Explore solutions through parameter adjustment and model modification to improve performance and accuracy, analyzing improvement effects and significance.

4.3 Results Analysis

4.3.1 Overall Analysis **Experimental Results**:
- **Model performance on test and validation sets**: The model achieved 92.7% accuracy on both test and validation sets across MGT, SNR0, and SNR5 datasets, indicating strong predictive capability and generalization on these datasets.
- **Model performance on new test set SNR1**: Accuracy dropped to 78.4% on the new SNR1 test set, indicating poor predictive capability and generalization.
- **Temporal prediction capability**: The model's prediction deviation for magnetic signal time points ranged between 30-50 seconds, indicating insufficient temporal prediction precision with potential lag or lead phenomena.

Results Analysis:
- **Reasons for good performance on test and validation sets**: The SGT model effectively captures long-range dependencies in one-dimensional data, improving magnetic signal classification accuracy. Additionally, using cross-entropy loss with Adam optimizer, appropriate learning rate, and weight decay facilitated model convergence and generalization.
- **Reasons for poor performance on unseen SNR1 test set**: The new dataset differs from training and validation sets in SNR, feature distribution, and sample quantity, affecting model adaptability and robustness. The model may also suffer from overfitting or underfitting, reducing generalization on unseen data.
- **Reasons for insufficient temporal prediction capability**: Suboptimal model structure and parameter settings may lead to inaccurate and ineffective temporal feature extraction and representation. The model may not adequately consider dynamic changes and periodicity in magnetic signals, causing temporal prediction errors and biases.

Improvement Directions:
To improve performance on unseen test sets, we plan to use more data for training and validation to increase data coverage and diversity, thereby enhancing adaptability and robustness. Additionally, we plan to employ regularization techniques such as Dropout and Batch Normalization to reduce overfitting/underfitting risks and improve generalization capability.

To improve temporal prediction capability, we will optimize model structure and parameters, such as increasing model depth and width and adjusting learning rate and weight decay to improve temporal feature extraction precision and effectiveness. We also plan to incorporate time series analysis techniques in future experiments to enhance the model's ability to capture and model dynamic changes and periodicity in magnetic signals, thereby improving temporal prediction accuracy and stability.

4.3.2 Anomaly Results Analysis Anomaly Analysis 1: High False Alarm Rate Due to Low Confidence Threshold in Continuity Judgment

Problem Description: When performing anomaly detection on time series data using a continuity judgment method, a confidence threshold must be set. When consecutive data points have confidence exceeding this threshold, they are considered anomalous. However, due to an excessively low confidence threshold, many normal data points were misclassified as anomalous, causing high false alarm rates.

Problem Causes: The low confidence threshold may stem from:
1. High noise levels in the data itself, leading to unstable and inaccurate continuity judgment.
2. Non-uniform data distribution, causing continuity judgment to be maladaptive and unreasonable.
3. Inherent defects in the continuity judgment method, as confidence intervals vary across datasets, leading to suboptimal performance.

Problem Resolution: To address high false alarm rates caused by low confidence thresholds, we plan to:
1. Preprocess data to reduce noise and improve quality and reliability.
2. Stratify or group data to set different confidence thresholds based on data characteristics or distributions, improving adaptability and reasonableness.
3. Improve or optimize the continuity judgment method by introducing additional criteria or metrics to enhance robustness and optimality.

Anomaly Analysis 2: Full Alarm Due to Overfitting

Problem Description: During time series anomaly detection, model overfitting on training data led to poor generalization on test data, causing all data points to be classified as anomalous (full alarm).

Problem Causes: Model overfitting may result from:
1. Insufficient training data quantity, leading to inadequate and ineffective learning.
2. Low training data quality, interfering with and affecting learning capability.
3. Excessive model complexity, causing overly sensitive and detailed learning.

Problem Resolution: To address full alarm caused by overfitting, we plan to:
1. Increase training data quantity for more adequate and effective learning.

2. Improve training data quality for more accurate and stable learning.
3. Reduce model complexity for more moderate and balanced learning.

Anomaly Analysis 3: Large Errors Due to Inaccurate Point Reporting

Problem Description: Using a single-feature extraction method for magnetic prospecting time series anomaly detection requires feature extraction and selection, with anomaly determination based on feature values. However, magnetic prospecting inherently depends on multiple features, and single-feature models easily cause false alarms and inaccuracies. In such cases, predictions deviate from actual anomaly points, causing large errors.

Problem Causes: Inaccurate feature extraction may result from:
1. Inappropriate feature extraction methods that don't align with data characteristics and patterns.
2. Non-optimized feature extraction parameters that don't adapt to data variations and differences.
3. Insufficient feature extraction dimensions, leading to incomplete and inadequate features.

Problem Resolution: To address large errors from inaccurate feature extraction, we plan to:
1. Adopt more appropriate feature extraction methods that align with data characteristics and patterns.
2. Optimize feature extraction parameters to adapt to data variations and differences.
3. Increase feature extraction dimensions for more complete and comprehensive features.

4.3.3 Improvement Method Testing **Problem Cause Summary**: Three main problems occurred in magnetic prospecting time series anomaly detection:
1. High false alarm rates due to low confidence thresholds in continuity judgment, where many normal points were misclassified as anomalous, affecting accuracy and efficiency.
2. Full alarm due to overfitting, where all points were classified as anomalous, affecting reasonableness and credibility.
3. Large errors due to inaccurate point reporting, where single-feature methods caused deviations between predictions and actual anomaly points, affecting precision and stability.

Improvement Direction Categorization: Based on these three problems, we identified improvement directions and decided to implement changes in three aspects:
1. Parameter tuning: Adjust model/method parameters to improve performance and accuracy while reducing errors and biases.
2. Optimize or replace feature extraction: Select appropriate feature extraction methods or improve existing ones to enhance feature quality and effectiveness, increase feature dimensions and diversity, and improve consistency and correla-

tion with anomaly points.
3. Replace continuity judgment with orthogonal basis judgment: Use orthogonal basis as the judgment criterion instead of continuity judgment to improve stability and reasonableness, reduce misclassification and false alarm risks, and enhance precision and optimality.

Improvement Directions and Measures:
1. **Parameter Tuning**: Increased confidence threshold from 0.5 to 0.7, reduced learning rate from 0.001 to 0.0005, increased weight decay from 0.0001 to 0.0005, and added a regularization module using dropout (found to work best at 0.2 through experiments).
2. **Optimize or Replace Feature Extraction**: Changed from single mean change rate to multi-feature comprehensive extraction, including mean change rate, standard deviation, kurtosis, and skewness.
3. **Replace Continuity Judgment with Orthogonal Basis Judgment**: Replaced final continuity judgment with orthogonal basis judgment using orthogonal basis as the criterion.

Improvement Effects and Analysis:
1. **Parameter Tuning**: Parameter adjustments improved model performance and accuracy, reducing errors and biases. Specifically, false alarm rates decreased from 0.25-0.35 to 0.15-0.25. Validation accuracy improved slightly, but generalization on new validation sets remained poor, with temporal deviation still fluctuating between 30-50 seconds. Analysis:
- Increasing confidence threshold reduced misclassification of normal points, improving stability and reasonableness.
- Adjusted parameters may not be optimal, leaving room for further optimization through more detailed grid search or Bayesian optimization.
- Adjusted parameters may conflict with other hyperparameters or model structure, preventing significant performance improvement through tuning alone. Model structure and parameter settings should be optimized to improve temporal feature extraction precision.

2. **Feature Extraction Optimization**: Experimentally switching from mean change rate to another feature did not fundamentally improve poor generalization and even worsened prediction effectiveness. Specifically, switching from mean change rate to standard deviation increased false alarm rates from 0.15-0.25 to 0.2-0.3 and decreased accuracy from 0.85 to 0.65. Analysis:
   - Single features may not fully reflect time series complexity and diversity, resulting in insufficient information and discriminative power to effectively distinguish normal and anomalous data.

   - Single features may contain noise or outliers, reducing feature quality and stability and affecting credibility and accuracy.

   - Single features may have unclear or inconsistent relationships with anomaly points, making it difficult to capture anomaly characteristics and patterns.

3. **Replacing Continuity Judgment with Orthogonal Basis Judgment**: This method did not integrate well with preceding feature fusion, causing false alarm rates to rise to 0.35, prediction point offset of 40-65 seconds, and accuracy to drop from 0.85 to 0.59. Analysis:
   - Orthogonal basis judgment parameters were not optimized, causing calculations to be maladaptive to data variations and affecting stability and robustness.

   - Orthogonal basis judgment dimensions were insufficient, leading to incomplete representation and affecting precision and optimality.

4.3.4 Future Improvement Ideas

1. **Comprehensive Multi-Feature Design**: Introduce multiple types of features such as time-domain, frequency-domain, and statistical features, and comprehensively utilize this information to build more complex and comprehensive models. This improves robustness and adaptability to different time series data types.
2. **Stacking Module for Multi-Feature Integration**: Introduce a stacking module that takes prediction results from single models based on different features as input and generates more accurate final predictions through further learning and synthesis. The stacking module helps fully leverage the advantages of each model and improve overall performance.
3. **Cross-Validation and Optimization**: Employ cross-validation in experiments to ensure good generalization performance across different datasets. Further optimize by adjusting hyperparameters and model structure to achieve optimal anomaly detection performance.
4. **Result Interpretation and Explainability**: Conduct result interpretation and explainability analysis for the final model to ensure prediction reasonableness. This helps better understand and trust model outputs in practical applications.

---

5. Conclusion

This report explored the application of the SGT model in magnetic prospecting, particularly its performance on MGT, SNR0, and SNR5 datasets. Experimental results revealed issues including high false alarm rates and large prediction biases. Subsequent improvement experiments were designed to address insufficient predictive and generalization capabilities from three perspectives: parameter tuning, feature extraction optimization, and modifying continuity judgment. Parameter tuning achieved approximately 0.5% performance improvement, but feature extraction optimization and orthogonal basis judgment methods exhibited compatibility issues with the model. Through code review and logical reasoning, we identified that the problem lies in incompatibility between extracted features and the model. To use orthogonal basis instead of continuity judgment requires integrating multiple features, while the traditional SGT model only extracts single features, failing to meet orthogonal basis algorithm requirements and resulting in degraded prediction performance. Through comprehensive analysis of experimental results and SGT model structure, we propose a new improvement approach: introduce multiple types of features such as time-domain, frequency-domain, and statistical features, and comprehensively utilize feature information to build a more comprehensive SGT model. Introduce a stacking module that takes prediction results from single models based on different features as input and generates more accurate predictions through further learning and synthesis.

References

[1] W. Hu, F. Wang, Q. Yin and F. Zhang, "SGT: A Generalized Processing Model for 1-D Remote Sensing Signal Classification," in *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022, Art no. 8030905.

[2] 徐磊, 张志强, 林朋飞, 等. 磁异常检测方法研究现状及发展趋势 [J]. 数字海洋与水下攻防, 2022(001):005.

[3] Luzhao CHEN, Wanhua ZHU, Peilin WU, Chunjiao FEI, Guangyou FANG. Magnetic Anomaly Detection Algorithm Based on Fractal Features Geomagnetic Background[J]. *Journal of Electronics & Information Technology*, 2019, 41(2): 332-340.

[4] JIN H, GUO J, WANG H, et al. Magnetic Anomaly Detection and Localization Using Orthogonal Basis of Magnetic Tensor Contraction[J/OL]. *Transactions on Geoscience and Remote Sensing*, 2020, 58(8): 5944-5954.

[5] TANG Y, LIU Z, PAN M, et al. Detection of Magnetic Anomaly Signal Based on Information Entropy of Differential Signal[J/OL]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(4): 512-516.

[6] ZHANG K Y, HU M K, DU C P, et al. Detection of Magnetic Dipole Target Signals by Using Convolution Neural Network[C/OL]//2018 Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC). Xuzhou, China: IEEE, 2018: 1-3.

[7] 银鸿, 文轩, 杨生胜, 等. 基于磁异常检测的磁性运动目标识别方法研究 [J]. 仪器仪表学报, 2018, 39(3):7.

[8] 费春娇, 张群英, 吴佩霖, 等. 一种海洋磁异常检测噪声抑制算法 [J]. 电子与信息学报, 2018, 40(11):8.

[9] Sheinker A, Frumkis L, Ginzburg B, et al. Magnetic Anomaly Detection Using a Three-Axis Magnetometer[J]. *IEEE Transactions on Magnetics*, 2009, 45(1):160-167.

[10] Zhang X, Liu H, Wang Z, et al. Anomaly detection of complex magnetic measurements using structured Hankel low-rank modeling and singular value decomposition[J]. *The Review of scientific instruments*, 2022, 93(4):045107.

Note: Figure translations are in progress. See original paper for figures.