# 融合经验模态分解与改进时域 Transformer 的网络安全态势预测

孙隽丰[1,2]，李成海[1]，宋亚飞[1]，倪　鹏[3]
（1. 空军工程大学防空反导学院，西安，710051；2. 94994 部队，南京，210000；
3. 复杂航空系统仿真重点实验室，北京，100076）

**摘要**　针对网络安全态势预测任务复杂且真实环境下数据噪声较大等问题，提出一种融合经验模态分解与改进时域 Transformer 的网络安全态势预测方法，通过“分解-重构”方式使用完全自适应噪声集合经验模态分解方法对真实环境下网络安全态势数据进行去噪预处理；提出改进时域 Transformer，使用时域 Transformer 模块提取网络安全态势数据序列的时间深层全局特征，并提出 Attention Fusion 机制实现时序特征的自适应融合，以更加稳健的特征融合方式完成预测任务。实验结果表明，本文提出的方法相较其他方法在预测精度方面具有显著提高，其拟合优度决定系数达到 0.997 860，拟合效果较好。

**关键词**　网络安全态势预测；时间序列分解；Transformer；特征融合；注意力机制
**DOI**　10.3969/j.issn.2097-1915.2024.06.013
**中图分类号**　TP393　　**文献标志码**　A　　**文章编号**　2097-1915(2024)06-0104-09

# A Network Security Situation Prediction Based on Empirical Mode Decomposition and Improved Temporal Transformer

SUN Junfeng[1,2], LI Chenghai[1], SONG Yafei[1], NI Peng[3]
(1. Air Defense and Antimissile School, Air Force Engineering University, Xi'an 710051, China;
2. Unit 94994, Nanjing 210000, China; 3. Science and Technology on Complex Aviation Systems Simulation Laboratory, Beijing 100076, China)

**Abstract**　Aimed at the problems that the network security situation prediction task is complex, and high in noise of data in real environments, a network security situation prediction method is proposed based on empirical mode decomposition (EMD) and improved temporal Transformer (ITTransformer). The complete EEMD with adaptive noise (CEEMDAN) method is utilized for de-noising and pre-processing network security situation data in real environments through "decomposition-reconstruction". The paper proposes ITTransformer. The Temporal Transformer module is used to extract the time-depth global features from the network security situation data sequences. An Attention Fusion mechanism is proposed to realize the adaptive fusion of temporal features to complete the prediction task in a more robust feature fusion

**收稿日期**：2023-12-12
**基金项目**：国家自然科学基金（62002362，61703426）；陕西省高校科协青年人才托举计划（2019038）；中国陕西省创新能力支持计划（2020KJXX-065）
**作者简介**：孙隽丰（1995－），男，山东烟台人，硕士生，研究方向为网络安全态势预测。E-mail：jf_sun2023@163.com
**通信作者**：宋亚飞（1988－），男，河南汝州人，副教授，博士，研究方向为智能信息处理。E-mail：yafei_song@163.com

*引用格式：孙隽丰，李成海，宋亚飞，等. 融合经验模态分解与改进时域 Transformer 的网络安全态势预测[J]. 空军工程大学学报，2024，25(6)：104-112. SUN Junfeng, LI Chenghai, SONG Yafei, et al. A Network Security Situation Prediction Based on Empirical Mode Decomposition and Improved Temporal Transformer[J]. Journal of Air Force Engineering University, 2024, 25(6): 104-112.*

way. The experimental results show that the method proposed in this paper is superior in prediction accuracy to the other methods, and its coefficient of determination reaches 0.997 860, and the fitting efficiency is good.

**Key words**　network security situation prediction; time series decomposition; Transformer; feature fusion; attention mechanism

信息技术的快速发展和广泛应用已经让互联网成为了人们日常生活中不可或缺的一部分。网络空间已经成为继陆、海、空、天之外的第五空间，承载着人类越来越多的活动，成为人类社会发展中不可或缺的重要内容[1]。随着网民规模的不断扩大，网络环境越发复杂。网络安全态势预测技术应运而生，相较于传统预测技术，其独特之处在于可将其视为相对主动积极的防御体系[2]。网络安全态势预测的视角将不再偏居一隅，而是从更加宏观的角度来进行计算与评估。在分析目标网络时，为分析者提供更加宏观与直观的数据，以此来展现目标网络的安全性。常见的预测方法有时间序列预测方法[3]、灰色理论预测、回归分析等。但现实情况是，网络安全态势的变化并不是简单的线性过程，这是因为网络攻击往往充满了随机与偶然。在处理具有非线性关系时，以上方法效果并不理想，已逐渐不能满足网络安全态势预测的需求。基于神经网络(Neural Network)、马尔可夫链(Markov)[4]、支持向量机(SVM)[5-6]等理论的预测方式与模型陆续由各学者发现。其中的神经网络方法在网络安全态势预测领域被广泛应用，其属于人工智能领域，在具备优秀的函数拟合性和自学习能力的同时可以进行并行处理，且容错程度高，为数据分析与处理提供有力支撑。

神经网络在网络安全态势预测领域的应用一直是当前研究人员关注的焦点。在传统的预测方法中，Preethi 等[7]提出了一种基于稀疏自动编码器驱动的支持向量回归(support vector regression，SVR)的网络入侵预测深度学习模型。它是一种自学习框架和无监督学习算法，减少了维度和训练时间，有效提高了预测精度。与传统方法相比，神经网络能有效地逼近和拟合非线性时序数据，并产生良好的情景预测结果。Zhang 等[8]提出了一种基于 SA-SOA 优化的 BP 神经网络的网络安全态势预测算法。该算法通过人群搜索算法(seeker optimization algorithm，SOA)寻找拟合度最优的个体，得到最优权值和阈值，并将其分配给 BP 神经网络。同时，在 SOA 中引入了模拟退火算法(simulated annealing algorithm，SA)，解决了搜索后期快速陷入局部优化和收敛速度慢的问题，提高了算法的全局搜索能力。Zhu 等[9]提出了一种基于改进 WGAN 的网络安全态势预测方法。该方法以 Wasserstein 距离作为损失函数，并在损失函数中加入了不同的项，正确解决了 GAN 训练费力和梯度不稳定的问题。Ni 等[10]提出了一种基于时间深度学习的网络安全态势预测。该方法将注意力机制与循环网络相结合，学习隐藏的历史时间序列数据特征。之后，对隐藏特征进行分析，并通过预测层预测网络安全状态。实验证明了未来安全态势预测所提模型的有效性。Xie 等[11]提出了一种基于功能演化网络的网络安全态势预测云模型。这一概念通过将进化算法与功能网络融合，创建了一个进化功能网络模型。同时，建立了受安全状况要素不确定性影响的关系可靠性矩阵。随机近似算法处理并理解由多元非线性回归算法预测的云安全状况的各个方面。在复杂的云网络环境中，它成功地解决了动态不确定性问题，提高了安全情景预测的准确性。赵冬梅等[12]通过引入门控循环单元(gate recurrent unit，GRU)来降低样本特征的维度，以减少训练 Transformer 的代价和缓解 Transformer 过拟合问题，提出了一种基于 GRU-Transformer 的网络安全态势预测模型。实验结果表明，该方法有效提高了预测模型的准确性和泛化能力。

网络安全态势预测任务非常复杂，且真实环境下数据噪声较大。为解决此问题，同时充分利用网络安全态势值的时序特征信息，增强全局信息的提取能力，提出一种融合经验模态分解与改进时域 Transformer(improved temporal transformer，IT-Transformer)的网络安全态势预测方法，本文主要贡献如下：

1)采用真实环境下的网络安全数据构建数据集，利用完全自适应噪声集合经验模态分解方法对其进行降噪，提高了预测精度。

2)研究 Transformer 在网络安全态势预测任务中的应用，利用改进的时域 Transformer 对网络安全态势数据序列之间的长程时序信息进行建模，进一步提高了网络安全态势预测任务的精度。

3)通过 Attention Fusion 的特征融合方法，对时间特征自适应地赋予注意力权重，并通过对比实验验证了所提融合方法的有效性。

# 1 融合经验模态分解与改进时域 Transformer 的模型

## 1.1 完全自适应噪声集合经验模态分解

针对 EMD 算法分解信号存在模态混叠的问题，EEMD 和 CEEMD 分解算法通过在待分解信号中引入成对正负高斯白噪声，以减轻模态混叠现象。然而，这 2 种算法通过分解信号后，其本征模态分量中会产生一定程度的白噪声残留，对后续信号的分析和处理产生影响[13]。为了解决这些问题，Torres 等[14]提出了一种改进算法，即完全自适应噪声集合经验模态分解（complete EEMD with adaptive noise，CEEMDAN），又称为完全集合经验模态分解。

CEEMDAN 原理如图 1 所示。

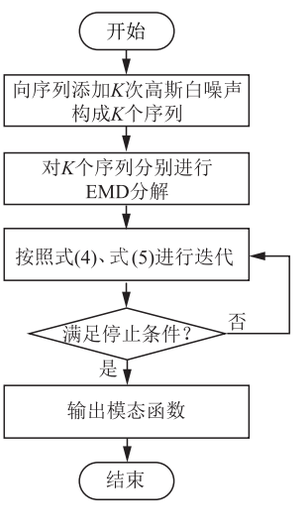


图 1 CEEMDAN 原理

Fig. 1 The schematic of CEEMDAN

CEEMDAN 分解过程如下：

1）构造共 $K$ 次实验的待分解序列 $s_i(t)$，$i=1,2,\cdots,K$，其中 $s_i(t)$ 为将待分解信号 $s(t)$ 添加 $K$ 次高斯白噪声（均值为 0）得到的结果。

$$s_i(t)=s(t)+\varepsilon\delta_i(t) \tag{1}$$

式中：$\varepsilon$ 为高斯白噪声权值系数；$\delta_i(t)$ 为第 $i$ 次处理时产生的高斯白噪声。

2）对上述序列 $s_i(t)$ 进行 EMD 分解，得到了第 1 个内涵模态分量（IMF），然后取第 1 个 IMF 的均值作为 CEEMDAN 分解得到的第 1 个 IMF。

$$\mathrm{IMF}_1(t)=\frac{1}{K}\sum_{i=1}^{K}\mathrm{IMF}_1^i(t) \tag{2}$$

$$r_1(t)=s(t)-\mathrm{IMF}_1(t) \tag{3}$$

式中：$\mathrm{IMF}_1(t)$ 为 CEEMDAN 分解得到的第 1 个内涵模态分量；$r_1(t)$ 为第 1 次分解后的余量信号。

3）将分解后得到的第 $j$ 阶段余量信号添加特定噪声后，继续进行 EMD 分解。

$$\mathrm{IMF}_j(t)=\frac{1}{K}\sum_{i=1}^{K}E_1(r_{j-1}(t)+\varepsilon_{j-1}E_{j-1}(\delta_i(t))) \tag{4}$$

$$r_j(t)=r_{j-1}(t)-\mathrm{IMF}_j(t) \tag{5}$$

式中：$\mathrm{IMF}_j(t)$ 为 CEEMDAN 分解得到的第 $j$ 个内涵模态分量；$E_{j-1}(\cdot)$ 为对序列进行 EMD 分解后的第 $j-1$ 个 IMF 分量；$\varepsilon_{j-1}$ 为 CEEMDAN 对第 $j-1$ 阶段余量信号加入噪声的权值系数；$r_j(t)$ 为第 $j$ 阶段余量信号。

4）迭代停止。如果满足第 $n$ 次分解的余量信号 $r_n(t)$ 为单调信号，则迭代停止，CEEMDAN 算法分解结束。

## 1.2 改进时域 Transformer

本文提出的改进时域 Transformer 的整体结构主要包括两部分，分别是时域 Transformer 模块和特征融合及预测模块，如图 2 所示。为适用于网络安全态势预测任务，仅使用 Transformer 框架的 Encoder 部分用于特征提取，通过堆叠 Encoder 提取深层时间的全局特征，而后将时间特征作为输入，经特征融合及预测模块进行特征融合和预测任务。

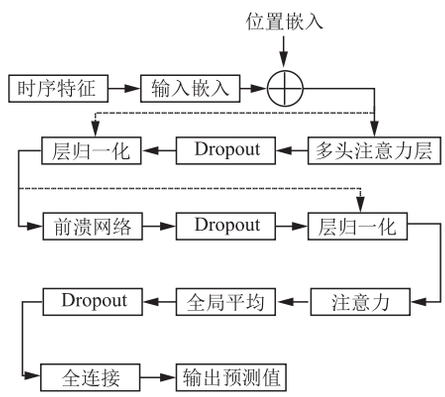


图 2 改进时域 Transformer 结构

Fig. 2 The structure of improved temporal Transformer

以序列 $\boldsymbol{S}=[\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_n]^{\mathrm{T}}$ 作为输入，序列的维度为 $n\times c$，其中 $n$ 为网络安全态势数据序列长度，$c$ 为序列中每个元素的通道维度。首先对序列内的元素经过高维线性映射从 $c$ 维处理为 $d_{\mathrm{model}}$ 维：

$$\boldsymbol{Y}_{\mathrm{token}}=\boldsymbol{S}\boldsymbol{W}_e \tag{6}$$

式中：$\boldsymbol{Y}_{\mathrm{token}}$ 为线性映射后的序列；$d_{\mathrm{model}}$ 为通道的维度；$\boldsymbol{W}_e\in\mathbb{R}^{c\times d_{\mathrm{model}}}$ 为线性映射的矩阵。

通过高维线性映射，能够将信息聚合到高维表示中，有利于自注意力机制进行特征提取。除此之外，通过位置编码，为 $\boldsymbol{Y}_{\mathrm{token}}$ 增加可学习的位置信息矩阵 $\boldsymbol{X}_p\in\mathbb{R}^{n\times d_{\mathrm{model}}}$，得到具有位置信息的高维序列 $\boldsymbol{Y}$。

以 $\boldsymbol{Y}$ 作为输入，由堆叠的 Encoder 进行特征提取，得到 Transformer 堆叠 $\boldsymbol{N}$ 个 Encoder 并输出深层时间全局特征 $\boldsymbol{O}_T$，通过 Attention Fusion 机制实现特征融合，特征权重的计算过程为：

$$\boldsymbol{A}_F=\mathrm{Softmax}(\boldsymbol{O}_T\boldsymbol{W}_F) \tag{7}$$

将时间特征 $\boldsymbol{A}_F$ 加权得到融合后的特征为：

$$\boldsymbol{O}_{\mathrm{Fusion}}=\boldsymbol{O}_T\boldsymbol{A}_F \tag{8}$$

融合特征包含丰富的时间特征，具有更强的能力，最后将融合特征作为输入，使用预测模块进行预测。

1.2.1　Encoder 结构

Encoder 是所提方法的重要组成部分，Encoder 主要由多头注意力模块、前馈网络（feed forward network，FFN）、层归一化和残差连接组成。多头注意力机制如图 3 所示。

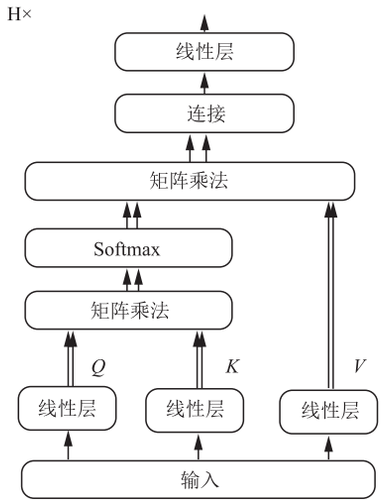


图 3　多头注意力机制示意

Fig. 3　Mechanism diagram of multi-head attention

将嵌入位置信息的高维序列 $\boldsymbol{Y}=[y_1,y_2,\ldots,y_n]$ 中的每个元素通过映射，得到查询向量 $\boldsymbol{Q}=\boldsymbol{Y}\boldsymbol{W}_Q$，键向量 $\boldsymbol{K}=\boldsymbol{Y}\boldsymbol{W}_K$ 和值向量 $\boldsymbol{V}=\boldsymbol{Y}\boldsymbol{W}_V$，其中 $\boldsymbol{W}_Q,\boldsymbol{W}_K,\boldsymbol{W}_V\in\mathbb{R}^{d_{\mathrm{model}}\times d_h}$，$d_h$ 为第 $h$ 个注意力头的维度。通过 $\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}$ 计算序列元素之间的注意力权值矩阵 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$：

$$\boldsymbol{A}=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_h}}\right) \tag{9}$$

使用注意力权值矩阵对值向量加权求得自注意力模块输出的加权序列值为：

$$\mathrm{SH}(\boldsymbol{X})=\boldsymbol{A}\boldsymbol{V} \tag{10}$$

多头注意力模块为 $H$ 个自注意力模块对输入序列进行处理，并将各个输出结果串行连接为：

$$\mathrm{MH}(\boldsymbol{X})=[\mathrm{SH}_1(\boldsymbol{X}),\mathrm{SH}_2(\boldsymbol{X}),\cdots,\mathrm{SH}_H(\boldsymbol{X})]W_{\mathrm{MH}} \tag{11}$$

式中：参数矩阵 $\boldsymbol{W}_{\mathrm{MH}}\in\mathbb{R}^{\mathrm{H}d_{\mathrm{H}}\times d_{\mathrm{model}}}$；$\mathrm{MH}(\boldsymbol{X})$ 为多头注意力模块的输出结果。

多头注意力模块将嵌入位置信息的高维向量组成的序列，通过可学习的参数矩阵分裂为查询向量、键向量和值向量，而后通过查询向量和键向量计算序列元素之间的相关性得到注意力权值矩阵，将值向量与注意力权值矩阵进行运算得到单头注意力模块的输出，通过合并单头注意力模块的输出得到多头注意力模块的输出结果。

将多头注意力模块的输出与输入完成残差连接和层归一化处理后，作为 FFN 的输入，FFN 包含 3 层，第 1 层先将输入映射到高维空间，第 2 层使用非线性激活层增强对特征的非线性表达能力，第 3 层再进行降维处理，过程为：

$$\boldsymbol{E}_l=(\mathrm{Relu}(\mathrm{LN}(\mathrm{MH}+\boldsymbol{Y})\boldsymbol{W}_1))\boldsymbol{W}_2 \tag{12}$$

$$\boldsymbol{O}_l=\mathrm{LN}(\boldsymbol{E}_l+(\mathrm{LN}(\mathrm{MH}+\boldsymbol{Y})) \tag{13}$$

式中：$\boldsymbol{O}_l$ 为第 $l$ 个 Encoder 的输出；$\boldsymbol{W}_1\in\mathbb{R}^{d_{\mathrm{model}}\times d_{\mathrm{FFN}}}$，$\boldsymbol{W}_2\in\mathbb{R}^{d_{\mathrm{FFN}}\times d_{\mathrm{model}}}$ 分别为两层维度变换的参数矩阵；LN(·)为层归一化函数。堆叠的 Encoder 为串行连接，将上一个 Encoder 的输出作为输入，最终输出深层全局特征。

1.2.2　时域 transformer 模块

网络安全态势数据序列由网络安全态势周报量化而成，随着时间的变化而动态改变，通过人工提取微小变化的深层时序信息难度较大。

为了提取网络安全态势数据序列之间的时序信息，以序列 $\boldsymbol{S}=[s_1,s_2,\ldots,s_n]$ 为输入，其中 $n$ 为序列的长度，$s_i$ 为第 $i$ 周的网络安全态势数据。Encoder 对序列之间的时间相关性进行建模，通过提取序列的动态时序变化信息，时域 Transformer 模块通过堆叠 $N$ 层 Encoder 挖掘网络安全态势数据序列深层时序全局特征，实现对序列的长程时序关系的有效表达。

1.2.3　预测模块

为了减少全连接层所需优化的参数，本文考虑通过特征融合得到的特征使用全局平均池化，来替代全部展开后输入到全连接层进行预测的方式。然而，由于需要调整超参数，不能完全依赖全局平均池化来替代全连接层。因此，本文在全连接层之前添加了一个全局平均池化层，以减少全连接层的参数数量，从而减缓过拟合问题。

除了全局平均池化外，本文还引入了 Dropout 机制，它在全局平均池化的基础上，以一定的概率选择性地忽略某些单元，以进一步缓解过拟合问题。

## 2　实验与分析

### 2.1　实验数据及预处理方法

为了验证本文所提出网络安全态势预测方法的有效性，本文使用国家互联网应急中心发布的网络安全信息与动态周报数据[15]作为实验基础。在数据集构建中，选取该网站自 2012 年第 30 期～2023

年第 35 期发布的共计 580 期周报数据作为基础来进行实验验证，这些数据主要从表 1 体现的 5 个角度对网络安全态势进行评估。为了更加直观地反映网络安全态势，本文采用文献[16]中提到的态势评估方法进行量化。根据网络安全威胁的程度不同，赋予其不同的权重，具体的权重分配如表 1 所示。然后，按照式(14)计算每周的态势值。

**表 1 网络安全威胁权重分配**

**Tab. 1 Weighting of network security threats**

| 威胁种类 | 权重 |
|---|---|
| 境内感染网络病毒的主机数量 | 0.30 |
| 境内被篡改网站总数 | 0.25 |
| 境内被植入后门网站总数 | 0.15 |
| 境内网站的仿冒页面数量 | 0.15 |
| 新增信息安全漏洞数量 | 0.15 |

$$\boldsymbol{S}=\sum_{i=1}^{5}\frac{\mathrm{NT}_i}{\mathrm{NT}_{\mathrm{imax}}}\omega_i \tag{14}$$

式中：$\mathrm{NT}_i$ 为周报中某种网络安全威胁的数量；$\mathrm{NT}_{\mathrm{imax}}$ 为选取的 580 期数据中该种安全威胁的最大数量；$\omega_i$ 为其对应的网络安全威胁权重。

2.1.1 数据归一化

数据归一化可以将特征的方差控制在一定范围内，减少异常值对模型性能的影响，从而提升模型的收敛速率。为了将特征数据规范至 $-1$～1，本文采用 min-max 归一化方法，其计算式为：

$$x'=\frac{x-\min(x)}{\max(x)-\min(x)} \tag{15}$$

式中：$x'$ 为 $x$ 映射到区间[$-1$，1]的数据；$\max(x)$和 $\min(x)$分别为数据集中数据的最大值和最小值。通过计算，得到数据归一化后的网络安全态势值。具体结果如图 4 所示。

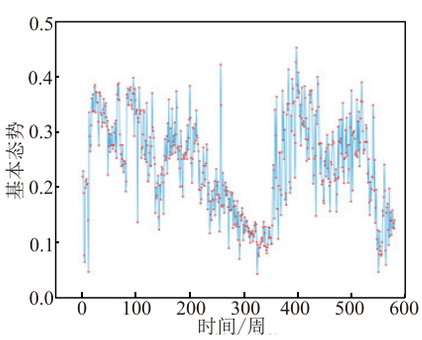


图 4 网络安全态势值

Fig. 4 Network security situation value

2.1.2 数据时序化

对经归一化处理后的网络安全态势值进行滑动窗口处理，将其转换为时间步长×输入维度的形式[17]。这种数据时序化的处理方式可以将当前数据和历史数据结合起来，用于预测未来的网络安全态势。以滑动窗口 $s=5$ 为例，此时设置的时间步长为 5，依次取 1～5、2～6 和 3～7 作为输入的样本，标签则分别为第 6、7、8 个样本的标签。数据重构如表 2 所示。

**表 2 数据重构结果**

**Tab. 2 Results of data reconstruction**

| 输入样本 | 输出样本 |
|---|---|
| $s_1,s_2,\cdots,s_5$ | $s_6$ |
| $s_2,s_3,\cdots,s_6$ | $s_7$ |
| ⋮ | ⋮ |
| $s_{251},s_{252},\cdots,s_{255}$ | $s_{256}$ |
| ⋮ | ⋮ |

## 2.2 评价指标与参数设置

为了评价提出的预测方法的效果，本文选择回归问题中常见的 4 个评价指标，即平均绝对误差(mean absolute error，MAE)、均方根误差(root mean square error，RMSE)、平均绝对百分比误差(mean absolute percentage error，MAPE)和拟合优度决定系数(the coefficient of determination)$R^2$。4 个评价指标的计算式分别为：

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i| \tag{16}$$

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{N}} \tag{17}$$

$$\mathrm{MAPE}=\frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i-y_i}{y_i}\right| \tag{18}$$

$$R^2=\frac{\left[\sum_{i=1}^{N}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}}_i)\right]^2}{\left[\sum_{i=1}^{N}(y_i-\bar{y})^2\right]\left[\sum_{i=1}^{N}(\hat{y}_i-\bar{\hat{y}}_i)^2\right]} \tag{19}$$

式中：$y_i$ 和 $\hat{y}_i$ 分别为某个样本的真实值和预测值；$N$ 为样本的个数；$\bar{y}_i$ 和 $\bar{\hat{y}}_i$ 分别为真实值的平均值和预测值的平均值。参数设置如表 3 所示。

**表 3 所提预测方法的具体参数设置**

**Tab. 3 Parameter settings for the proposed forecasting method**

| 参数 | 参数设置 |
|---|---|
| 损失函数 | MAE |
| 优化器 | Adam |
| 批处理大小 | 32 |
| 学习率 | 0.0001 |
| 迭代轮次 | 200 |
| Encoder 的数量 | 6 |
| 多头注意力的头数 | 4 |
| Embedding 的维度 | 8 |
| 隐含层单元数 | 64 |
| Dropout | 0.2 |

## 2.3 数据分解

实验对数据集进行分解预处理，将计算后的网络安全态势值使用 CEEMDAN 分解后得到的结果如图 5 所示。

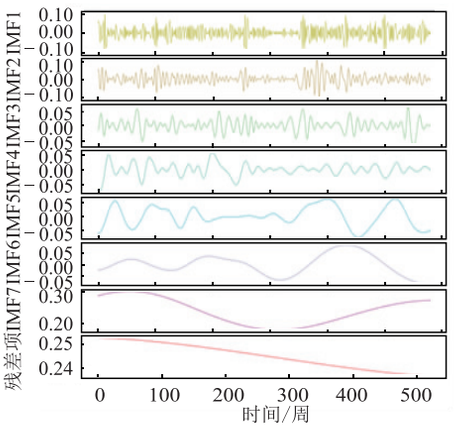


图 5　经 CEEMDAN 分解后的结果

Fig. 5　Results after CEEMDAN decomposition

由图 5 可知，使用 CEEMDAN 将原网络安全态势数据序列分解为 7 个 IMF 和 1 个残差项，前 2 个内涵模态分量 IMF1 和 IMF2 为高频分量，幅度在(−0.1，0.1)之间。本文通过去除部分高频 IMF，再对剩余分量采用叠加重构的方法实现对数据的去噪。经过实验验证，选择重构参数为 2 时，对序列进行重构去噪的效果最佳。

### 2.4　实验结果与分析

为验证本文所提网络安全态势预测方法的性能，本文设置多组实验。

#### 2.4.1　不同分解方式分析

将本文所用分解方式与不使用分解、使用 EMD 和使用 EEMD 进行对比实验，各项评价指标对比结果如表 4 所示。

**表 4　不同分解方式评价指标对比**

**Tab. 4　Comparison of evaluation indicators for different decompositions**

| 分解方式 | MAE | RMSE | MAPE/% | $R^2$ |
|---|---|---|---|---|
| IT Transformer | 0.016 383<br>(−0.013 135) | 0.022 488<br>(−0.018 696) | 9.064 100<br>(−7.096 719) | 0.924 748<br>(+0.073 112) |
| EMD-IT Transformer | 0.012 151<br>(−0.008 903) | 0.015 817<br>(−0.012 025) | 6.706 691<br>(−4.739 310) | 0.962 772<br>(+0.035 088) |
| EEMD-IT Transformer | 0.008 981<br>(−0.005 733) | 0.011 211<br>(−0.007 418) | 4.593 043<br>(−2.625 662) | 0.981 299<br>(+0.016 561) |
| CEEMDAN-IT Transformer | 0.003 248<br>(0.000 000) | 0.003 792<br>(0.000 000) | 1.967 381<br>(0.000 000) | 0.997 860<br>(0.000 000) |

注：括号内数值为本文所提模型相较其他模型评价指标的绝对变化量

实验结果表明，经 CEEMDAN 分解后再使用 IT-Transformer 比仅使用 ITTransformer 方法 MAE 降低了 80.17%，RMSE 降低了 83.14%，MAPE 降低了 78.29%，$R^2$ 提高了 7.91%；比经 EMD 分解 MAE 降低了 73.27%，RMSE 降低了 76.02%，MAPE 降低了 70.67%，$R^2$ 提高了 3.64%；比经 EEMD 分解 MAE 降低了 63.83%，RMSE 降低了 66.17%，MAPE 降低了 57.17%，$R^2$ 提高了 1.69%。实验证明了 CEEMDAN 分解的有效性，不同分解方式预测的网络安全态势值如图 6 所示。

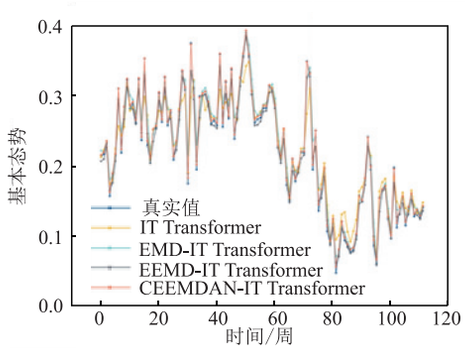


图 6　不同分解方式预测的网络安全态势值

Fig. 6　Network security situation values predicted by different decompositions

#### 2.4.2　不同池化方式分析

将本文所用池化方式与使用全局最大池化进行对比实验，各项评价指标对比结果如表 5 所示。

**表 5　不同池化方式评价指标对比**

**Tab. 5　Comparison of evaluation indicators for different pooling methods**

| 池化方式 | MAE | RMSE | MAPE/% | $R^2$ |
|---|---|---|---|---|
| 全局最大池化 | 0.014 401<br>(−0.011 153) | 0.018 326<br>(−0.014 534) | 9.852 396<br>(−7.885 015) | 0.950 027<br>(+0.047 833) |
| 全局平均池化 | 0.003 248<br>(0.000 000) | 0.003 792<br>(0.000 000) | 1.967 381<br>(0.000 000) | 0.997 860<br>(0.000 000) |

注：括号内数值为本文所提模型相较其他模型评价指标的绝对变化量

实验结果表明，使用全局平均池化比使用全局最大池化 MAE 降低了 77.45%，RMSE 降低了 79.31%，MAPE 降低了 80.03%，$R^2$ 提高了 5.03%。通过实验可以看出，选用全局平均池化可以使预测方法精度更高，不同池化方式预测的网络安全态势值如图 7 所示。

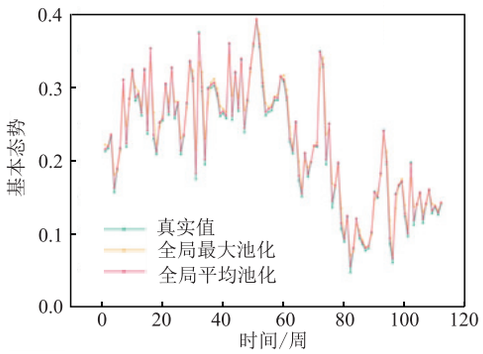


图 7 不同池化方式预测的网络安全态势值

Fig. 7 Network security situation values predicted by different pooling methods

2.4.3 不同特征融合方式分析

在 Transformer 框架下的特征融合任务中，主要使用 Add Fusion 和 Concatenate Fusion 的融合方式，为证明本文所提特征融合方式的有效性，与 Add Fusion 和 Concatenate Fusion 进行对比实验，各项评价指标对比结果如表 6 所示。

**表 6 不同特征融合方式评价指标对比**

**Tab. 6 Comparison of evaluation indicators for different feature fusion methods**

| 特征融合方式 | MAE | RMSE | MAPE/% | $R^2$ |
|---|---|---|---|---|
| Add Fusion | 0.012 916<br>(−0.009 668) | 0.016 934<br>(−0.013 141) | 7.707 323<br>(−5.729 942) | 0.957 331<br>(+0.040 529) |
| Concatenate Fusion | 0.016 629<br>(−0.013 381) | 0.020 757<br>(−0.016 965) | 8.225 682<br>(−6.258 301) | 0.935 887<br>(+0.061 973) |
| Attention Fusion | 0.003 248<br>(0.000 000) | 0.003 792<br>(0.000 000) | 1.967 381<br>(0.000 000) | 0.997 860<br>(0.000 000) |

注：括号内数值为本文所提模型相较其他模型评价指标的绝对变化量

实验结果表明，本文所提特征融合方式比 Add Fusion 特征融合方式 MAE 降低了 74.85%，RMSE 降低了 77.60%，MAPE 降低了 74.47%，$R^2$ 提高了 4.23%；比 Concatenate Fusion 特征融合方式 MAE 降低了 80.47%，RMSE 降低了 81.73%，MAPE 降低了 76.08%，$R^2$ 提高了 6.62%。实验发现，仅通过相加或者同维度连接进行特征融合，而不考虑特征之间的相关性是不合适的，本文所提特征融合方式通过注意力机制学习时序特征的相关性，对特征赋予权重，实现更加有效的自适应融合。不同特征融合方式预测的网络安全态势值如图 8 所示。

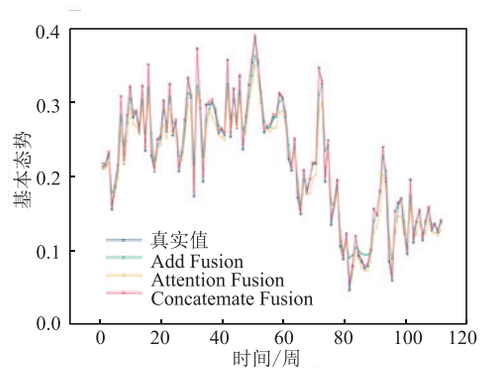


图 8 不同特征融合方式预测的网络安全态势值

Fig. 8 Network security situation values predicted by different feature fusion approaches

2.4.4 不同方法评价指标分析

为验证所提方法的有效性，选择 BP、LSTM、GRU、TCN、Transformer 5 种方法作为基线方法进行对比实验。各种方法的各项评价指标对比结果如表 7 所示。

**表 7 不同方法评价指标对比**

**Tab. 6 Comparison of evaluation indicators for different methods**

| 方法 | MAE | RMSE | MAPE/% | $R^2$ |
|---|---|---|---|---|
| BP | 0.039 978<br>(−0.036 730) | 0.051 421<br>(−0.047 629) | 22.514 054<br>(−20.546 673) | 0.610 019<br>(+0.387 841) |
| LSTM | 0.037 964<br>(−0.034 716) | 0.048 176<br>(−0.044 383) | 22.061 537<br>(−20.094 156) | 0.654 650<br>(+0.343 210) |
| GRU | 0.035 487<br>(−0.032 239) | 0.045 078<br>(−0.041 286) | 18.713 316<br>(−16.745 935) | 0.697 633<br>(+0.300 227) |
| TCN | 0.025 996<br>(−0.022 748) | 0.033 472<br>(−0.029 680) | 14.689 000<br>(−12.721 619) | 0.833 288<br>(+0.164 571) |
| Transformer | 0.023 734<br>(−0.020 486) | 0.030 191<br>(−0.026 399) | 12.893 368<br>(−10.925 987) | 0.864 370<br>(+0.133 490) |
| CEEMDAN-IT Transformer | 0.003 248<br>(0.000 000) | 0.003 792<br>(0.000 000) | 1.967 381<br>(0.000 000) | 0.997 860<br>(0.000 000) |

注：括号内数值为本文所提模型相较其他模型评价指标的绝对变化量

实验结果表明，本文所提方法比 BP 方法 MAE 降低了 91.88%，RMSE 降低了 92.62%，MAPE 降低了 91.26%，$R^2$ 提高了 63.58%；比 LSTM 方法 MAE 降低了 91.44%，RMSE 降低了 92.13%，MAPE 降低了 91.08%，$R^2$ 提高了 52.43%；比 GRU 方法 MAE 降低了 90.85%，RMSE 降低了 91.59%，MAPE 降低了 89.49%，$R^2$ 提高了 43.04%；比 TCN 方法 MAE 降低了 87.51%，RMSE 降低了 88.67%，MAPE 降低了 86.61%，$R^2$ 提高了 19.75%；比 Transformer 方法 MAE 降低了 86.32%，RMSE 降低了 87.44%，MAPE 降低了 84.74%，$R^2$ 提高了 15.44%。通过实验结果分析，证明了本文所提方法的有效性，不同方法预测的网络安全态势值如图 9 所示。

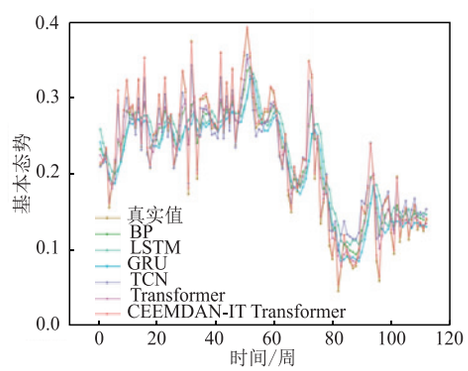


图 9　不同方法预测的网络安全态势值

Fig. 9　Network security situation values predicted by different methods

### 2.4.5　不同方法收敛性分析

图 10 为不同方法的损失值随迭代轮次增加的变化曲线。从图中可以观察到，本文提出的方法在收敛速度和收敛精度方面都优于其他模型。这证明了本文所提出的模型能够有效地学习网络安全态势的时序特征，并取得较好的效果。

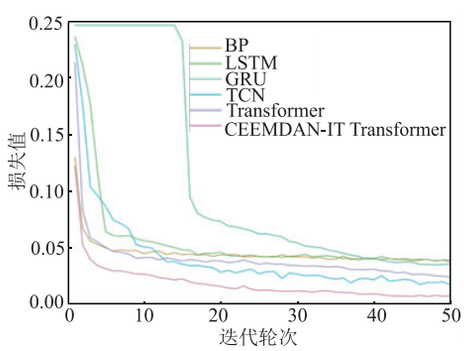


图 10　不同方法的损失值随迭代轮次增加的变化曲线

Fig. 10　Curves of loss values for different methods with increasing iteration rounds

## 3　结语

本文提出一种融合经验模态分解与改进时域 Transformer 的网络安全态势预测方法，通过使用完全自适应噪声集合经验模态分解方法对真实环境下网络安全态势数据进行去噪预处理，使用时域 Transformer 模块提取网络安全态势数据序列的时间深层全局特征，并将时序特征通过 Attention Fusion 机制实现自适应融合。在网络安全态势预测数据集的实验表明，所提方法的精确性有显著提高，收敛速度提升较大。但本文模型对不同类型、不同长度的网络安全风险或攻击处理能力较弱，在现有工作的基础上，下一步将针对不同类型、不同长度的网络攻击以及网络安全态势的多步预测进行研究，同时进一步提高 Transformer 在小样本上的预测性能。

## 参考文献

[1]　张玉清，刘奇旭，付安民，等. 新时代的网络空间治理[J]. 信息安全研究，2021，7(6)：486-487.
ZHANG Y Q，LIU Q X，FU A M，et al. Cyberspace Governance in the New Era[J]. Journal of Information Security Research，2021，7(6)：486-487. (in Chinese)

[2]　JAJODIA S，LIU P，SWARUP V，et al. Cyber Situational Awareness[M]. Springer US，2010.

[3]　BOX G E P，JENKINS G M，REINSEL G C. Time Series Analysis Forecasting and Control[M]. Beijing：Posts & Telecom Press，2005：19-180.

[4]　LIANG W，CHEN Z，YAN X L，et al. Multiscale Entropy-Based Weighted Hidden Markov Network Security Situation Prediction Model[C]//2017 IEEE International Congress on Internet of Things (ICIOT). Honolulu，HI：IEEE，2017：97-104.

[5]　ZHANG S M，LI B X，WANG B Y. The Application of an Improved Integration Algorithm of Support Vector Machine to the Prediction of Network Security Situation[J]. Applied Mechanics and Materials，2014，513：2285-2288.

[6]　DUAN M. Short-Time Prediction of Traffic Flow Based on PSO Optimized SVM[C]//2018 International Conference on Intelligent Transportation，Big Data & Smart City (ICITBS). Xiamen：IEEE，2018：41-45.

[7]　PREETHI D，KHARE N. Sparse Auto Encoder Driven Support Vector Regression Based Deep Learning Model for Predicting Network Intrusions[J]. Peer-to-Peer Networking and Applications，2021，14(4)：2419-2429.

[8]　ZHANG R，LIU M，YIN Y F，et al. Prediction Algorithm for Network Security Situation Based on BP Neural Network Optimized by SA-SOA[J]. International Journal of Performability Engineering，2020，16(8)：1171.

[9] ZHU J, WANG T T. Network Security Situation Prediction Based on Improved WGAN [M]. Cham: Springer International Publishing, 2019: 654-664.

[10] NI W D, GUO N W. Network Security Situation Prediction with Temporal Deep Learning[C]//2020 International Conference on Image, Video Processing and Artificial Intelligence. Shanghai: SPIE, 2020, 11584: 475-480.

[11] XIE B W, ZHAO G S, CHAO M X, et al. A Prediction Model of Cloud Security Situation Based on Evolutionary Functional Network[J]. Peer-to-Peer Networking and Applications, 2020, 13(5): 1312-1326.

[12] 赵冬梅，李志坚. 基于 Transformer 的网络安全态势预测[J]. 华中科技大学学报(自然科学版)，2022，50(5)：46-52.
ZHAO D M, LI Z J. Network Security Situation Prediction Based on Transformer[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2022, 50(5): 46-52. (in Chinese)

[13] 魏炘，石强，符文熹，等. 考虑 CEEMDAN 样本熵和 SVR 的短期风速预测[J]. 水电能源科学，2020，38(11)：207-210.
WEI X, SHI Q, FU W X, et al. Short-Term Wind Speed Prediction with CEEMDAN Sample Entropy and SVR[J]. Water Resources and Power, 2020, 38(11): 207-210. (in Chinese)

[14] TORRES M E, COLOMINAS M A, SCHLOTTHAUER G, et al. A Complete Ensemble Empirical Mode Decomposition with Adaptive NOISE[C]//2011 IEEE Imternational Conference on Acoustics, Speech and Signal Processing (ICASSP). Prague: IEEE, 2011: 4144-4147.

[15] 中国国家互联网应急中心. 2012 至 2023 年网络安全信息与动态周报[EB/OL]. [2023-08-31]. https://www.cert.org.cn/publish/main/index.html.
CNCERT/CC. Weekly Report on Network Security Information from 2012 to 2023[EB/OL]. [2023-08-31]. https://www.cert.org.cn/publish/main/index.html. (in Chinese)

[16] 姜万菲. 基于多模型权重提取与融合的网络安全态势预测研究[D]. 兰州：兰州理工大学，2016.
JIANG W F. Research on Network Security Situation Prediction based on Multi-Model Weight Extraction and Fusion [D]. Lanzhou: Lanzhou University of Technology, 2016. (in Chinese)

[17] 孙隽丰，李成海，曹波. 基于 TCN-BiLSTM 的网络安全态势预测[J]. 系统工程与电子技术，2023，45(11)：3671-3679.
SUN J F, LI C H, CAO B. Network Security Situation Prediction Based on TCN-BiLSTM[J]. Systems Engineering and Electronics, 2023, 45(11): 3671-3679. (in Chinese)

(编辑：杜娟)