

SEER：基于自动Patch增强与替换的鲁棒性时间序列预测模型研究报告

摘要 (Abstract)

本报告针对现实世界时间序列数据中普遍存在的低质量问题（如缺失值、分布偏移、异常值和白噪声）进行了深入研究。现有基于Patch的方法通常利用所有Patch进行建模，未能有效过滤包含低质量信息的Patch，从而导致预测性能下降。本研究介绍了一种名为 **SEER** 的新型Transformer框架，通过“增强嵌入模块”和“可学习的Patch替换模块”来解决上述问题。实验结果表明，SEER在多种低质量数据场景下表现出了卓越的鲁棒性，并在多个基准数据集上取得了最先进（SOTA）的预测精度。

1. 引言与背景 (Introduction & Background)

1.1 研究背景

时间序列预测在经济、交通、医疗和能源等领域至关重要。近年来，基于Patch（分块）的方法（如PatchTST）通过将时间序列划分为具有语义意义的片段，显著提高了捕捉长程依赖的能力。

1.2 问题陈述

然而，现实世界的数据收集过程面临诸多挑战，导致数据质量下降：

- 缺失值 (Missing Values)**: 传感器故障或传输中断导致。
- 分布偏移 (Distribution Shift)**: 数据生成机制随时间演变。
- 异常值 (Anomalies)**: 系统故障或突发事件引入。
- 白噪声 (White Noise)**: 传感器采集过程中不可避免的噪声。

这些因素导致部分Patch包含无效或负面信息，如果模型不加筛选地使用这些Patch，会导致性能恶化。

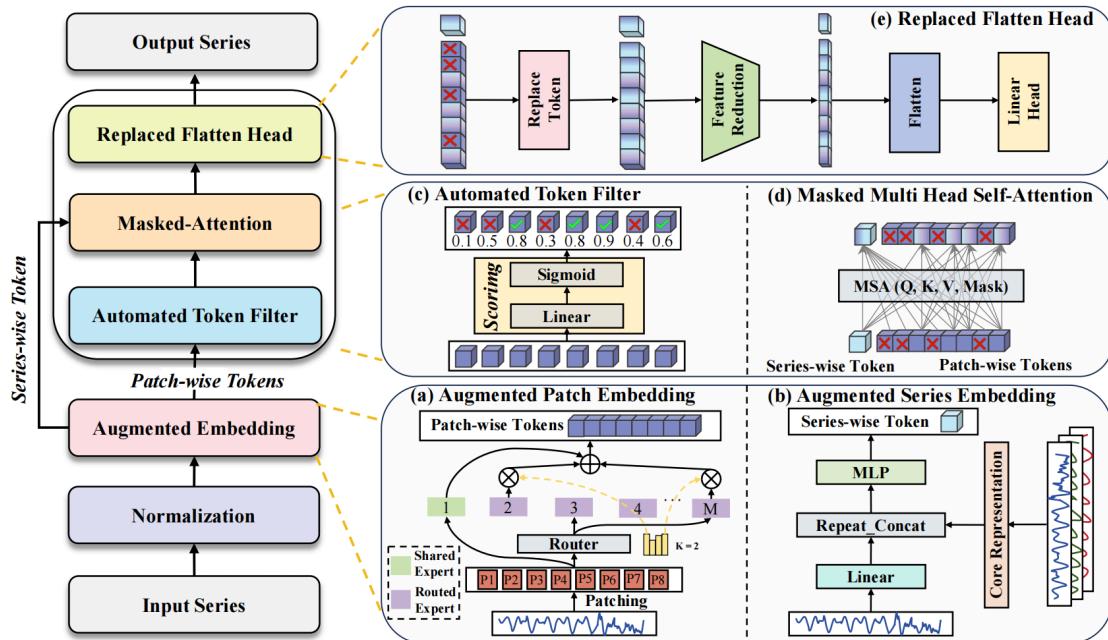
1.3 本文贡献

本研究提出了SEER框架，核心贡献包括：

- 设计了增强嵌入模块 (Augmented Embedding Module)，利用混合专家 (MoE) 架构增强Patch表示。
- 提出了可学习的Patch替换模块 (Learnable Patch Replacement Module)，动态剔除负面Patch并利用全局优化的序列级Token填补语义空白。

2. 方法论 (Methodology)

SEER框架旨在通过双尺度 (Patch级和Series级) 表示增强以及动态替换机制来提高算法的准确性和鲁棒性。



2.1 增强嵌入模块 (Augmented Embedding Module)

该模块旨在解决非鲁棒时间序列中表示能力不足的问题。

- 增强Patch嵌入 (Augmented Patch Embedding):** 并没有使用单一的线性投影，而是采用了 混合专家 (Mixture-of-Experts, MoE) 架构。通过路由机制 (Routing Mechanism) 自适应地聚合具有相似模式的Patch，构建多个异构线性表示空间，从而丰富Patch的语义特征。
- 增强序列嵌入 (Augmented Series Embedding):** 利用通道自适应感知机制 (Channel-adaptive perception mechanism) 来捕获时间序列的全局时序模式，生成序列级Token (Series-wise Token) 作为后续替换的“原型”。

2.2 可学习的Patch替换模块 (Learnable Patch Replacement Module)

这是SEER提升鲁棒性的核心，包含三个阶段：

1. 自动Token过滤 (Automated Token Filter)：

- 通过一个评分机制 (Scoring Mechanism) 评估每个Patch的质量。
- 动态识别并剔除含有噪声或无信息的负面Patch。
- 公式化表示为利用Sigmoid函数生成分数，并根据阈值 τ 生成保留或过滤的掩码。

2. 掩码多头自注意力 (Masked Multi Head Self-Attention)：

- 在保留的Patch Token和全局Series Token之间建立依赖关系。
- 利用掩码机制隔离已被过滤的Token，仅让“好”的Token与全局Token进行交互，从而优化全局表示。

3. 替换Flatten头 (Replaced Flatten Head)：

- 使用经过全局优化的Series Token来替换掉第一步中被过滤掉的Patch Token。
- 这填补了过滤操作留下的语义空白，最后通过自适应降维投影压缩关键信息并进行预测。

3. 实验设置 (Experimental Settings)

3.1 数据集

实验在9个知名的预测基准数据集上进行，涵盖多个领域：

- ETT (4个子集)
- Weather, Exchange, Electricity, Solar, Traffic

3.2 基线模型 (Baselines)

对比了10个基线模型，包括2025年的最新SOTA模型：

- 2025 SOTA: DUET, TimeKAN, xPatch, TimePro, Amplifier
- 经典/近期SOTA: iTransformer, PatchTST, TimesNet, DLinear

3.3 评估指标

采用 **均方误差 (MSE)** 和 **平均绝对误差 (MAE)** 作为评估指标。

4. 结果与分析 (Results & Analysis)

4.1 主实验结果

- SEER在绝大多数情况下表现优异。在45个MSE设置中获得了33个第一，在35个MAE设置中获得了42个第一。
- 与第二好的基线模型 (DUET) 相比，SEER的MSE平均降低了4.2%，MAE平均降低了3.5%。
- 即便在数据质量较低的数据集 (如Exchange和ETT) 上，SEER依然保持了良好的性能，验证了替换模块的有效性。

Models	SEER (ours)		DUET (2025)		TimeKAN (2025)		xPatch (2025)		TimePro (2025)		Amplifier (2025)		Fredformer (2024)		iTransformer (2024)		PatchTST (2023)		TimesNet (2023)		DLinear (2023)		
Metrics	mse	mae	mse	mae	mse	mae	mse	mae	mse	mae	mse	mae	mse	mae	mse	mae	mse	mae	mse	mae	mse	mae	
ETH1	96	0.365	0.384	0.377	0.393	0.374	0.391	0.378	0.390	0.375	0.398	0.376	0.393	0.378	0.395	0.386	0.405	0.414	0.419	0.384	0.402	0.397	0.412
	192	0.414	0.416	0.429	0.425	0.421	0.433	0.420	0.427	0.429	0.442	0.430	0.435	0.424	0.441	0.436	0.460	0.445	0.436	0.429	0.446	0.441	
	336	0.449	0.437	0.471	0.446	0.464	0.440	0.484	0.445	0.472	0.450	0.478	0.446	0.485	0.447	0.487	0.458	0.501	0.466	0.491	0.469	0.489	0.467
	720	0.451	0.457	0.496	0.480	0.466	0.462	0.480	0.476	0.474	0.501	0.479	0.499	0.472	0.503	0.491	0.500	0.488	0.521	0.500	0.513	0.510	
	avg	0.420	0.424	0.443	0.436	0.431	0.429	0.444	0.429	0.438	0.449	0.437	0.448	0.435	0.454	0.448	0.469	0.455	0.458	0.450	0.461	0.458	
ETH2	96	0.282	0.330	0.296	0.345	0.293	0.343	0.287	0.332	0.293	0.345	0.298	0.347	0.291	0.342	0.297	0.349	0.302	0.348	0.340	0.374	0.340	0.394
	192	0.358	0.377	0.368	0.389	0.375	0.396	0.360	0.382	0.367	0.394	0.378	0.401	0.372	0.390	0.380	0.400	0.388	0.400	0.402	0.414	0.482	0.479
	336	0.401	0.413	0.411	0.422	0.429	0.411	0.417	0.421	0.419	0.431	0.428	0.437	0.419	0.431	0.428	0.432	0.426	0.433	0.452	0.452	0.591	0.541
	720	0.388	0.418	0.412	0.434	0.466	0.468	0.412	0.433	0.427	0.445	0.452	0.460	0.431	0.450	0.427	0.445	0.431	0.446	0.462	0.468	0.839	0.661
	avg	0.357	0.385	0.372	0.397	0.391	0.412	0.369	0.392	0.377	0.403	0.389	0.411	0.378	0.403	0.387	0.407	0.414	0.427	0.563	0.519		
ETTm1	96	0.308	0.338	0.324	0.354	0.327	0.365	0.316	0.343	0.326	0.364	0.318	0.356	0.326	0.361	0.334	0.368	0.329	0.367	0.338	0.375	0.346	0.374
	192	0.362	0.369	0.369	0.379	0.363	0.387	0.369	0.369	0.367	0.383	0.362	0.381	0.363	0.384	0.377	0.391	0.367	0.385	0.374	0.387	0.382	0.391
	336	0.394	0.389	0.404	0.402	0.389	0.407	0.401	0.392	0.402	0.409	0.393	0.404	0.399	0.406	0.426	0.420	0.399	0.410	0.410	0.411	0.415	0.415
	720	0.458	0.432	0.463	0.437	0.457	0.445	0.461	0.429	0.469	0.446	0.460	0.442	0.456	0.441	0.491	0.459	0.454	0.430	0.478	0.450	0.473	0.451
	avg	0.381	0.382	0.390	0.393	0.384	0.401	0.387	0.383	0.391	0.400	0.383	0.396	0.385	0.398	0.407	0.410	0.387	0.400	0.400	0.406	0.404	0.408
ETTm2	96	0.171	0.250	0.174	0.255	0.178	0.262	0.174	0.252	0.178	0.260	0.178	0.261	0.177	0.258	0.180	0.264	0.175	0.259	0.187	0.267	0.193	0.293
	192	0.238	0.294	0.243	0.302	0.244	0.308	0.240	0.297	0.242	0.303	0.243	0.303	0.243	0.301	0.250	0.309	0.241	0.302	0.249	0.309	0.284	0.361
	336	0.296	0.331	0.304	0.341	0.305	0.346	0.302	0.335	0.303	0.342	0.305	0.344	0.302	0.340	0.311	0.348	0.305	0.343	0.321	0.351	0.382	0.429
	720	0.396	0.390	0.399	0.397	0.402	0.404	0.403	0.393	0.400	0.399	0.393	0.397	0.404	0.398	0.412	0.407	0.402	0.400	0.408	0.403	0.558	0.525
	avg	0.275	0.316	0.280	0.324	0.282	0.330	0.280	0.319	0.281	0.326	0.280	0.326	0.281	0.324	0.288	0.332	0.281	0.326	0.291	0.333	0.354	0.402
Weather	96	0.162	0.198	0.163	0.202	0.164	0.210	0.166	0.202	0.166	0.207	0.165	0.210	0.163	0.207	0.174	0.214	0.177	0.218	0.172	0.220	0.195	0.252
	192	0.211	0.242	0.218	0.252	0.205	0.250	0.210	0.242	0.216	0.254	0.212	0.253	0.224	0.258	0.221	0.254	0.225	0.259	0.219	0.261	0.237	0.295
	336	0.268	0.284	0.274	0.294	0.264	0.290	0.267	0.284	0.273	0.296	0.267	0.293	0.278	0.298	0.278	0.296	0.279	0.280	0.306	0.282	0.331	
	720	0.344	0.336	0.349	0.343	0.343	0.342	0.344	0.335	0.351	0.346	0.344	0.342	0.357	0.350	0.358	0.349	0.354	0.348	0.365	0.345	0.382	
	avg	0.246	0.265	0.251	0.273	0.245	0.273	0.247	0.266	0.251	0.276	0.247	0.275	0.256	0.278	0.258	0.278	0.259	0.281	0.259	0.287	0.265	0.315
Solar	96	0.185	0.201	0.200	0.207	0.228	0.254	0.234	0.245	0.196	0.237	0.186	0.232	0.189	0.236	0.203	0.237	0.234	0.286	0.250	0.292	0.290	0.378
	192	0.220	0.222	0.228	0.233	0.241	0.284	0.265	0.262	0.231	0.263	0.231	0.264	0.224	0.258	0.221	0.254	0.225	0.259	0.219	0.261	0.320	0.398
	336	0.237	0.239	0.262	0.244	0.273	0.292	0.301	0.280	0.250	0.281	0.234	0.263	0.243	0.286	0.248	0.273	0.290	0.315	0.319	0.330	0.353	0.415
	720	0.245	0.241	0.258	0.279	0.279	0.308	0.284	0.253	0.285	0.265	0.285	0.250	0.285	0.249	0.275	0.289	0.317	0.338	0.337	0.357	0.413	
	avg	0.222	0.226	0.237	0.233	0.232	0.266	0.277	0.268	0.232	0.266	0.222	0.256	0.227	0.267	0.233	0.262	0.270	0.307	0.301	0.319	0.330	0.401
Exchange	96	0.080	0.198	0.086	0.205	0.083	0.202	0.081	0.198	0.085	0.204	0.080	0.207	0.084	0.202	0.086	0.206	0.088	0.205	0.107	0.234	0.088	0.218
	192	0.169	0.291	0.182	0.305	0.179	0.301	0.175	0.296	0.178	0.299	0.175	0.299	0.172	0.299	0.177	0.299	0.176	0.299	0.226	0.344	0.176	0.315
	336	0.319	0.408	0.310	0.403	0.331	0.417	0.342	0.422	0.328	0.414	0.356	0.433	0.316	0.406	0.331	0.417	0.301	0.397	0.367	0.448	0.313	0.427
	720	0.681	0.621	0.693	0.624	0.843	0.692	0.817	0.679	0.888	0.711	0.850	0.693	0.847	0.691	0.901	0.714	0.964	0.746	0.839	0.695		
	avg	0.312	0.380	0.318	0.384	0.322	0.386	0.360	0.402	0.352	0.399	0.377	0.412	0.355	0.398	0.360	0.403	0.367	0.404	0.416	0.443	0.354	0.414
Electricity	96	0.142	0.225	0.145	0.233	0.174	0.266	0.160	0.244	0.139	0.234	0.149	0.245	0.148	0.242	0.148	0.240	0.195	0.285	0.168	0.272	0.210	0.302
	192	0.158	0.240	0.163	0.248	0.182	0.273	0.169	0.253	0.156	0.249	0.165	0.260	0.165	0.257	0.162	0.253	0.199	0.285	0.184	0.289	0.210	0.305
	336	0.174	0.258	0.175	0.262	0.197	0.286	0.185	0.268	0.172	0.267	0.176	0.271	0.180	0.274	0.178	0.269	0.215	0.305	0.198	0.300	0.223	0.319
	720	0.199	0.279	0.204	0.291	0.236	0.320	0.221	0.300	0.209	0.299	0.204	0.296	0.218	0.305	0.225	0.317	0.226	0.337	0.220	0.320	0.258	0.350
	avg	0.168	0.250	0.172	0.2																		

- 结果：**如下图表所示，SEER在所有四种低质量数据条件下均表现出最小的性能下降，优于DLinear、TimesNet等增强了鲁棒性插件的模型。
- 特别是在分布偏移场景下，SEER表现出了极强的泛化能力。

Models	Missing Values		Distribution Shift		Anomalies		White Noise	
Metrics	mse	mae	mse	mae	mse	mae	mse	mae
DLinear	0.170	0.264	0.179	0.273	0.182	0.277	0.166	0.259
DUET	0.166	0.256	0.164	0.259	0.179	0.264	0.172	0.256
TimesNet	0.198	0.284	0.202	0.285	0.208	0.292	0.198	0.283
iTransformer	0.181	0.273	0.187	0.279	0.188	0.278	0.180	0.273
SEER	0.165	0.250	0.167	0.254	0.173	0.259	0.163	0.252

4.3 消融实验 (Ablation Studies)

对SEER的各个组件进行了移除测试，以验证其贡献：

- w/o MoE:** 性能在复杂数据集（如ETTh2）上下降，证明MoE对增强复杂模式表示的重要性。
- w/o Token Filter:** 性能显著下降，证明过滤噪声Token对维持精度至关重要。
- w/o Replace Token:** 如果不使用全局优化的Token进行替换，性能大幅下降，证明了“填补语义空白”步骤的必要性。

Datasets	ETTh2		ETTm2		Weather		Solar	
Metrics	mse	mae	mse	mae	mse	mae	mse	mae
w/o Mixture-of-Experts	0.361	0.390	0.280	0.322	0.249	0.267	0.228	0.232
w/o Token Filter	0.366	0.390	0.288	0.326	0.250	0.269	0.228	0.231
w/o Feature Reduction	0.368	0.392	0.277	0.318	0.248	0.267	0.226	0.229
w/o Replace Token	0.365	0.389	0.278	0.318	0.249	0.268	0.227	0.228
SEER	0.357	0.385	0.275	0.316	0.246	0.265	0.222	0.226

4.4 参数敏感性分析 (Parameter Sensitivity Analysis)

为了验证关键超参数对模型性能的影响，研究针对以下四个方面进行了敏感性分析：

1. 专家数量 (Number of Experts):

- 观察：**在ETTh2数据集上，随着MoE中专家数量的增加，模型性能呈现提升趋势。

- **结论：**专家数量的增加有助于模型捕捉更复杂的模式，这对于像 ETTh2 这种非平稳数据集尤为重要。

2. Patch 长度 (Patch Length):

- **观察：**大多数数据集在 Patch 长度设置为 **16 或 24** 时达到最佳性能。
- **结论：**Patch 长度过小或过大都会导致性能下降，适中的长度最有利于特征提取。

3. 缩放比例 (Scaling Ratio):

- **观察：**在特征降维 (Feature Reduction) 模块，缩放比例设置为 **0.5 或 0.75**，效果最佳。
- **结论：**这表明特征压缩不仅能减少模型参数，还能一定程度通过压缩信息来提升预测性能。

4. 评分阈值 (Score Threshold):

- **观察：**在自动 Token 过滤器模块中，当评分阈值设置为**非零值**，大多数数据集的表现最好。
- **结论：**这一结果证实了适度丢弃那些具有负面影响的 Token（即噪声或无效信息）确实能增强模型的整体表现。

5. 结论 (Conclusion)

SEER框架通过引入MoE增强的嵌入层和创新的“过滤-交互-替换”机制，成功解决了低质量时间序列数据对预测模型的干扰。该模型不仅在标准基准测试中达到了SOTA水平，更在面对缺失、噪声和漂移数据时展现了卓越的鲁棒性，为高可靠性的时间序列预测提供了新的解决方案。

10244511415 沈天恩