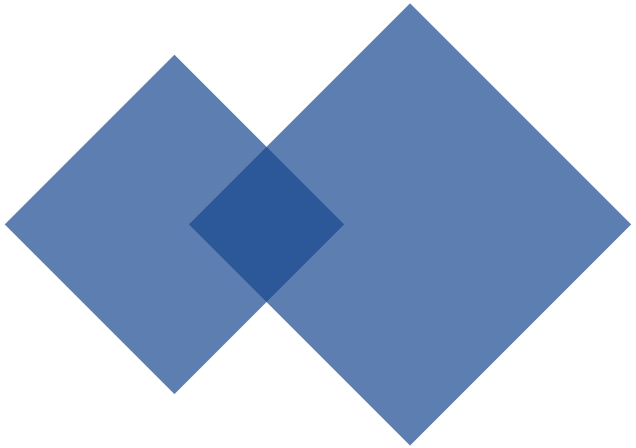
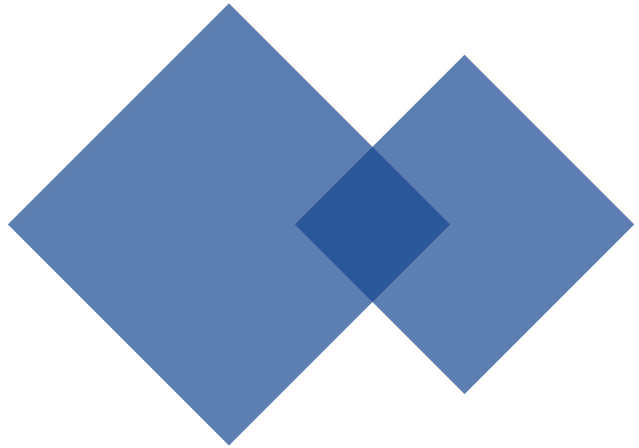


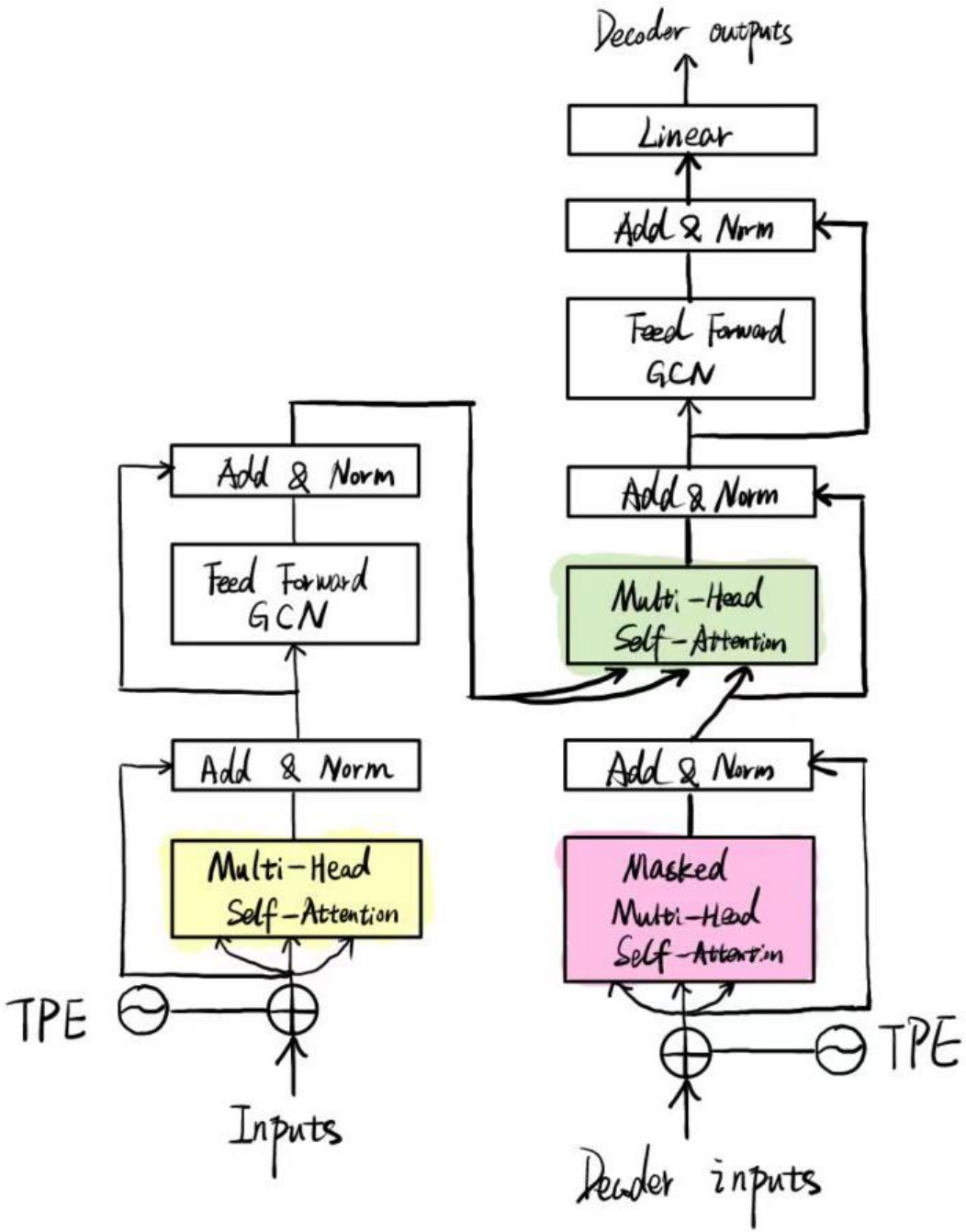
GBT

Two-stage transformer framework for non-stationary time series forecasting



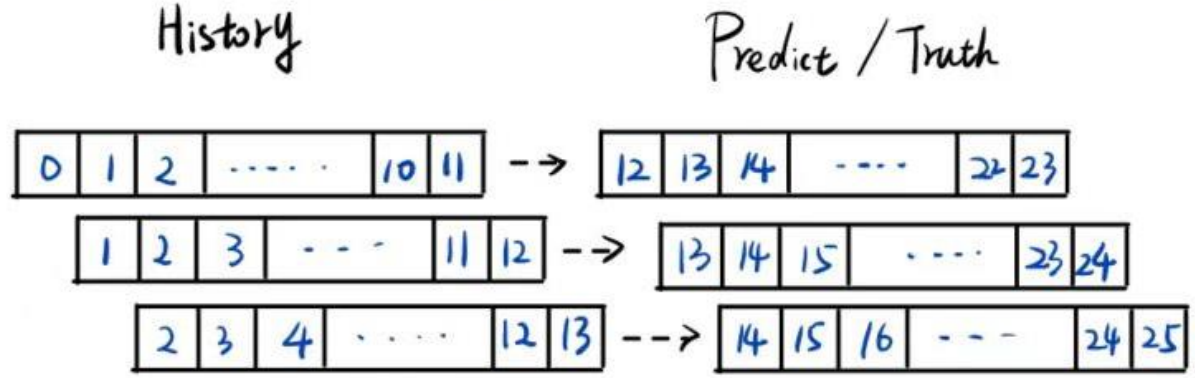
23.12.12

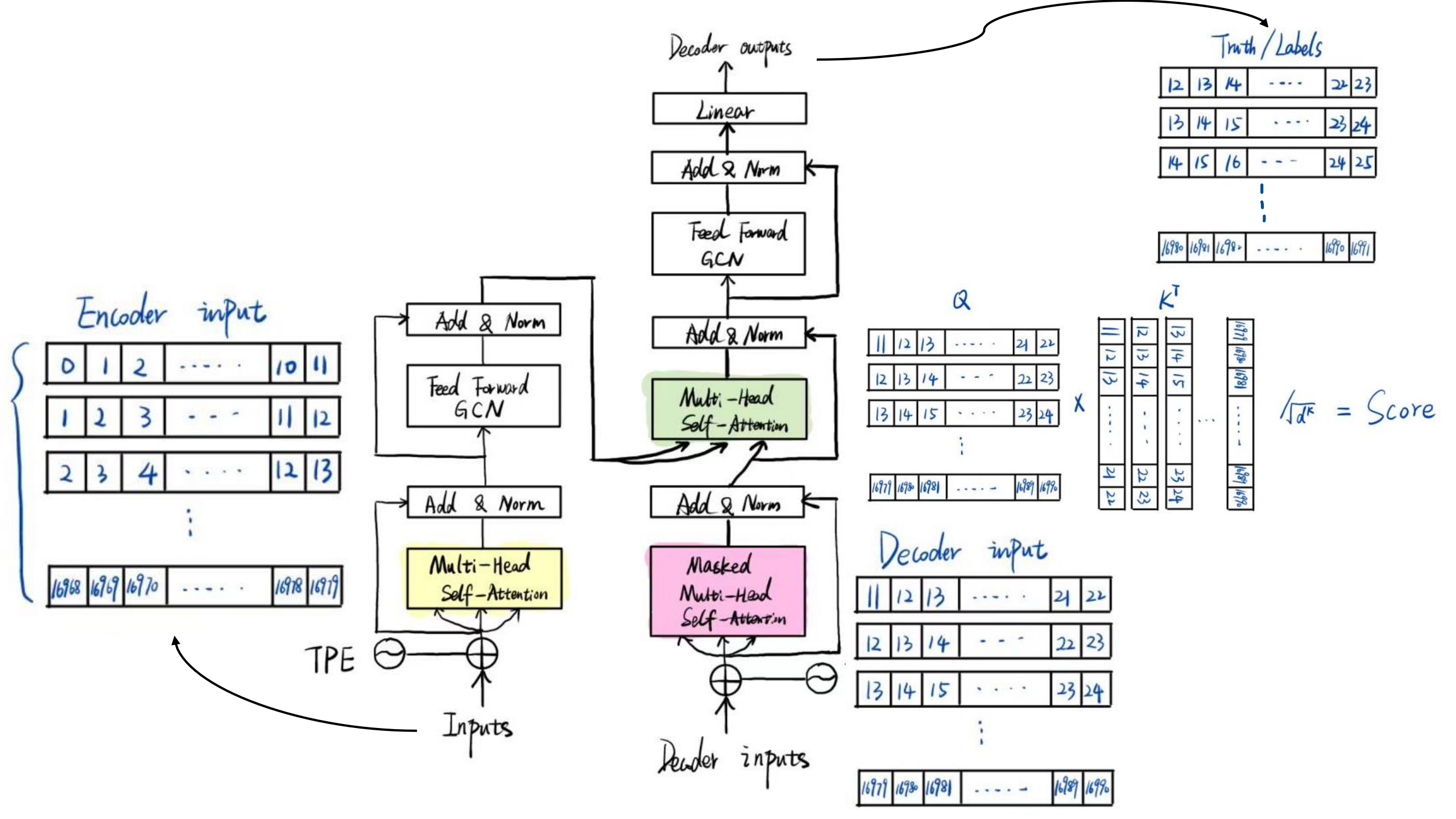
Presented by Yyyq

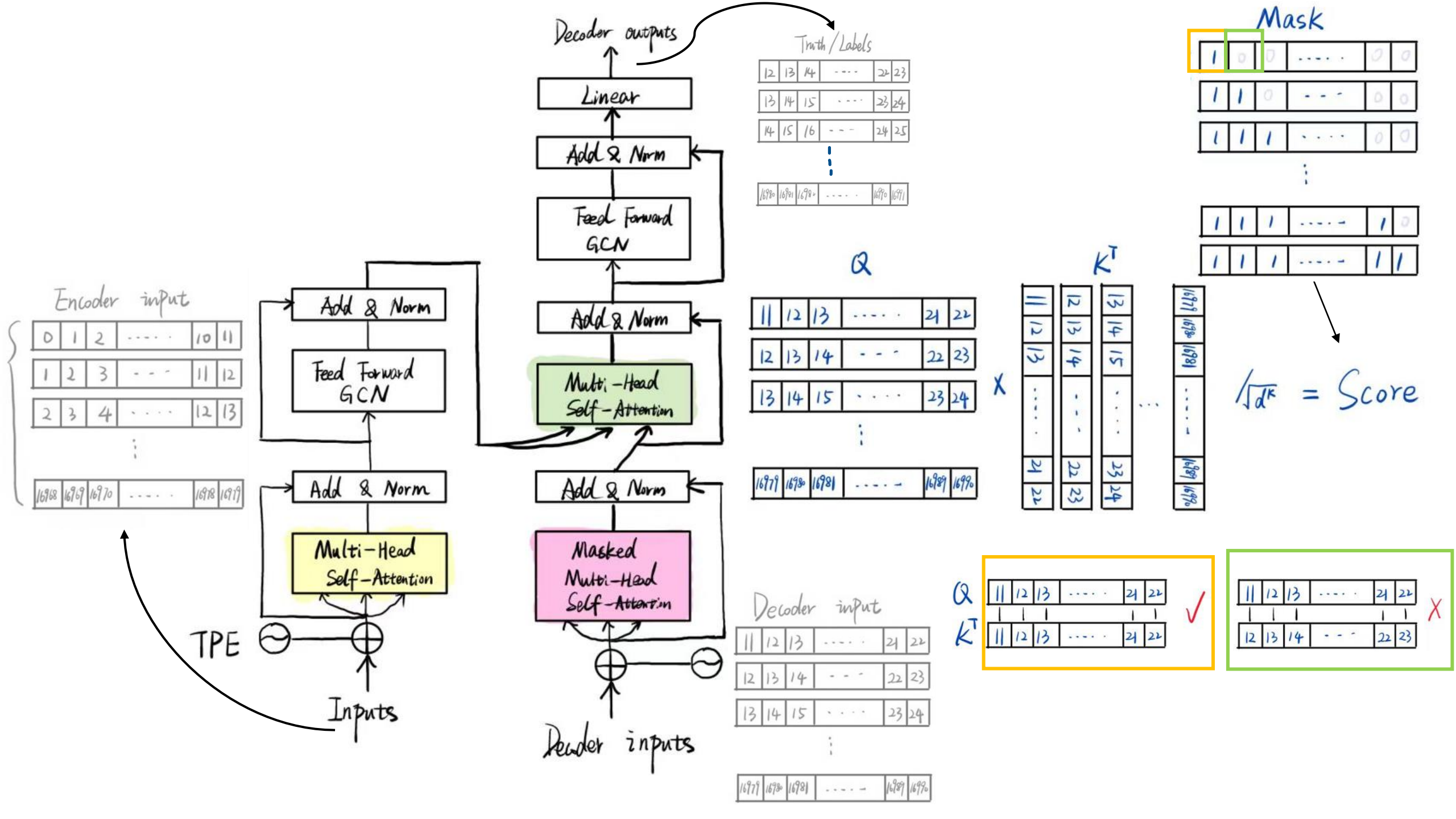


59x288:59天×(1小时12点六天24小时)

PEMSD4 (16992, 307, 3)

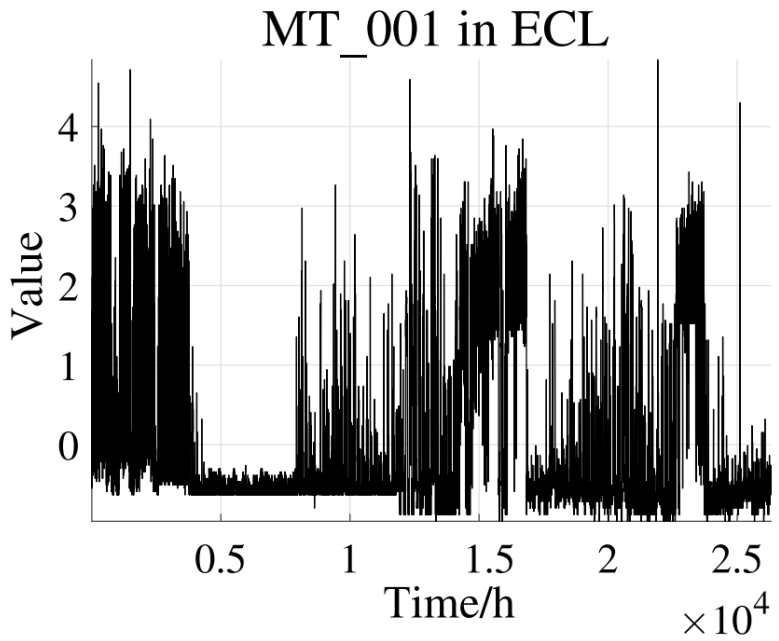



Encoder input
Add & Norm
Feed Forward GCN
Add & Norm
Multi-Head Self-Attention
TPE
Inputs
Decoder outputs
Linear
Add & Norm
Feed Forward GCN
Add & Norm
Multi-Head Self-Attention
Add & Norm
Masked Multi-Head Self-Attention
Decoder inputs
Truth / Labels
Q
K^T
X
√d^K = Score
Decoder input


Encoder input
Add & Norm
Feed Forward GCN
Add & Norm
Multi-Head Self-Attention
Add & Norm
TPE
Inputs
Decoder outputs
Linear
Add & Norm
Feed Forward GCN
Add & Norm
Multi-Head Self-Attention
Add & Norm
Masked Multi-Head Self-Attention
Decoder inputs
Truth / Labels
Decoder input
Q
K^T
Mask
√d^k = Score

01 问题描述

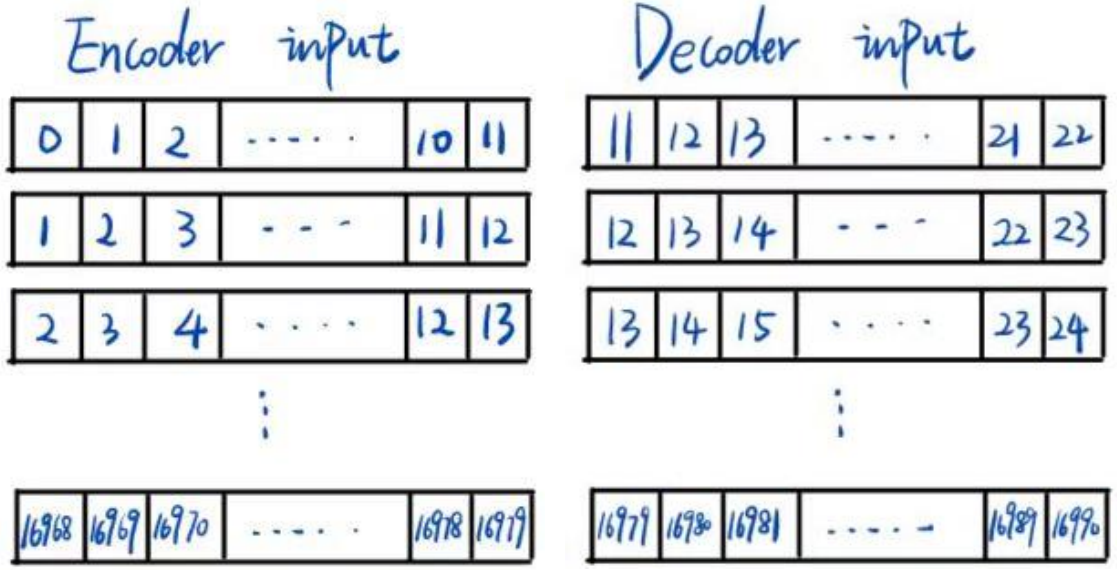
- **Time Series Forecasting Transformer (TSFT)**
- **非平稳时间序列：** （平稳性）
  - 一个平稳的时间序列意味着它的统计性质在时间上是不变的，即均值和方差不随时间而变化。
  - 平稳的时间序列应该在统计性质上不随时间变化，因此图形上看起来应该没有明显的趋势。



01 问题描述

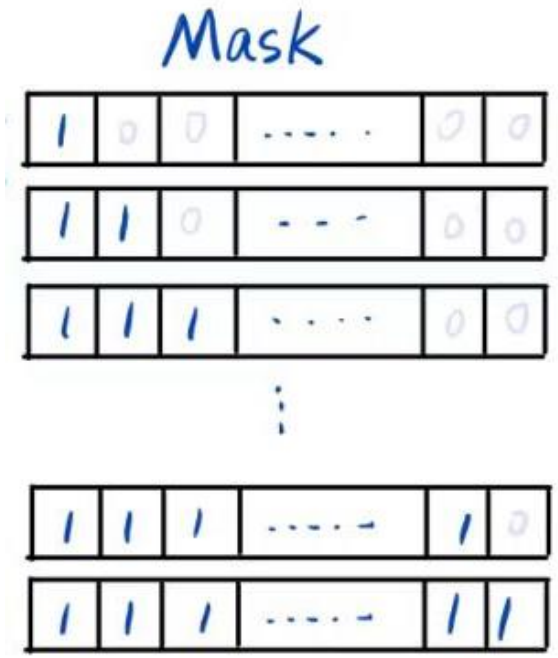
- **TSFT处理非平稳时间序列的能力较弱**
- **存在过拟合问题：**
  - 不同局部窗口的统计动态不同（输入/输出序列的数据分布不同）
  - 最近工作表明了MLP模型的先进表现，说明TSFT存在不足

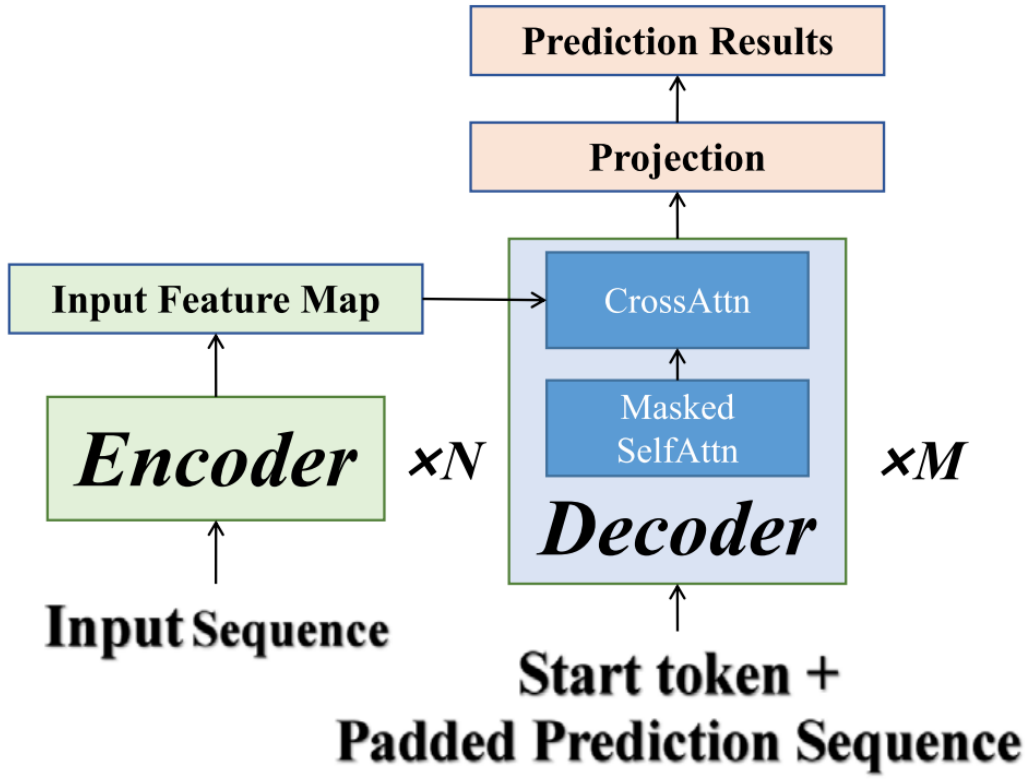
Decoder Inputs 的初始化方法存在问题



01 问题描述

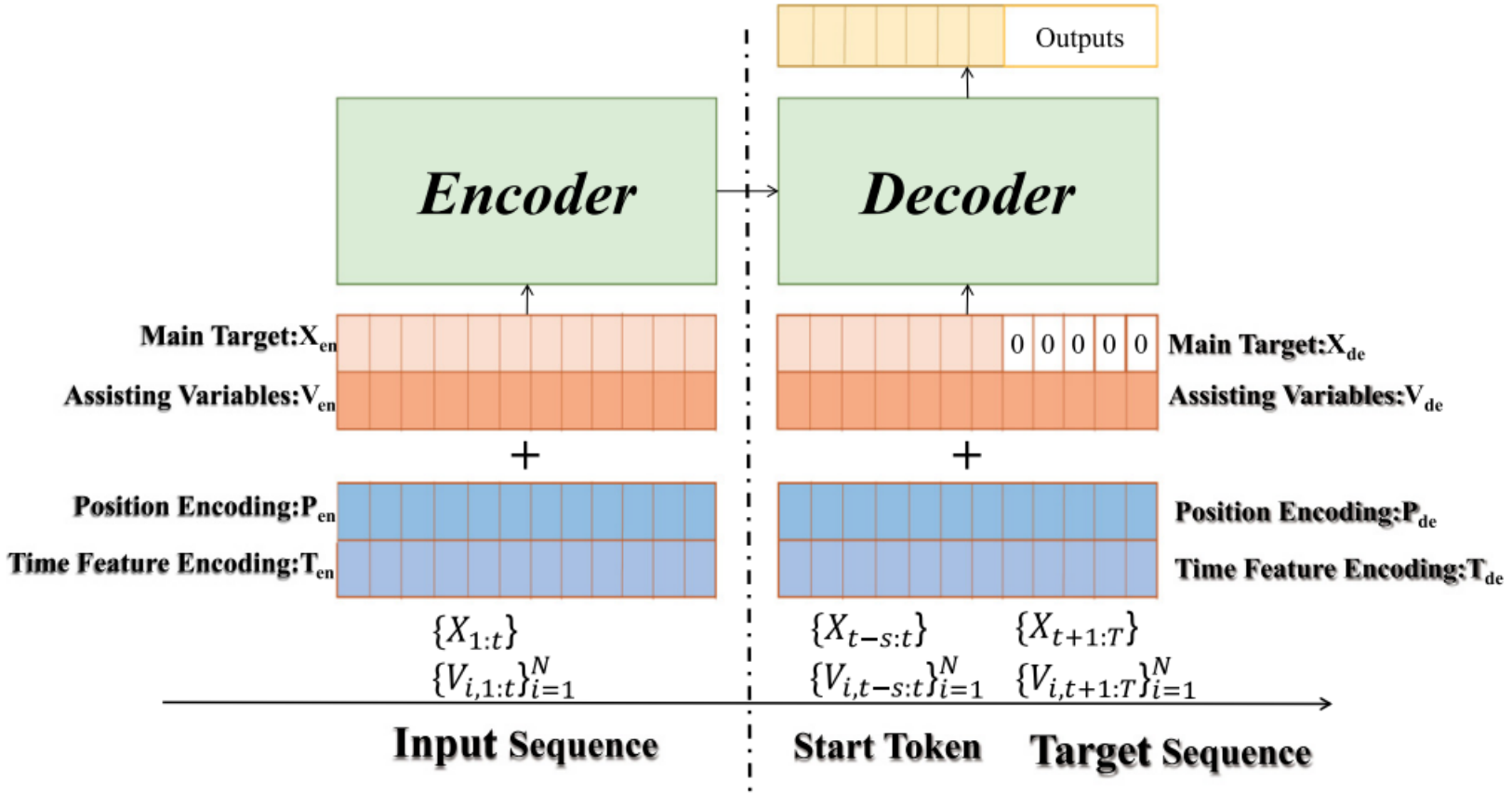
- **Decoder Inputs 的初始化方法存在问题**
  - Start token：Informer
  - 趋势分解方法：FEDformer





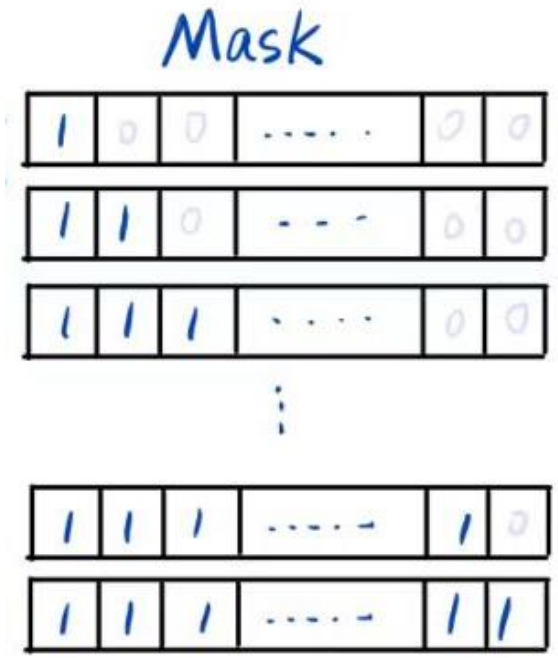
01 问题描述

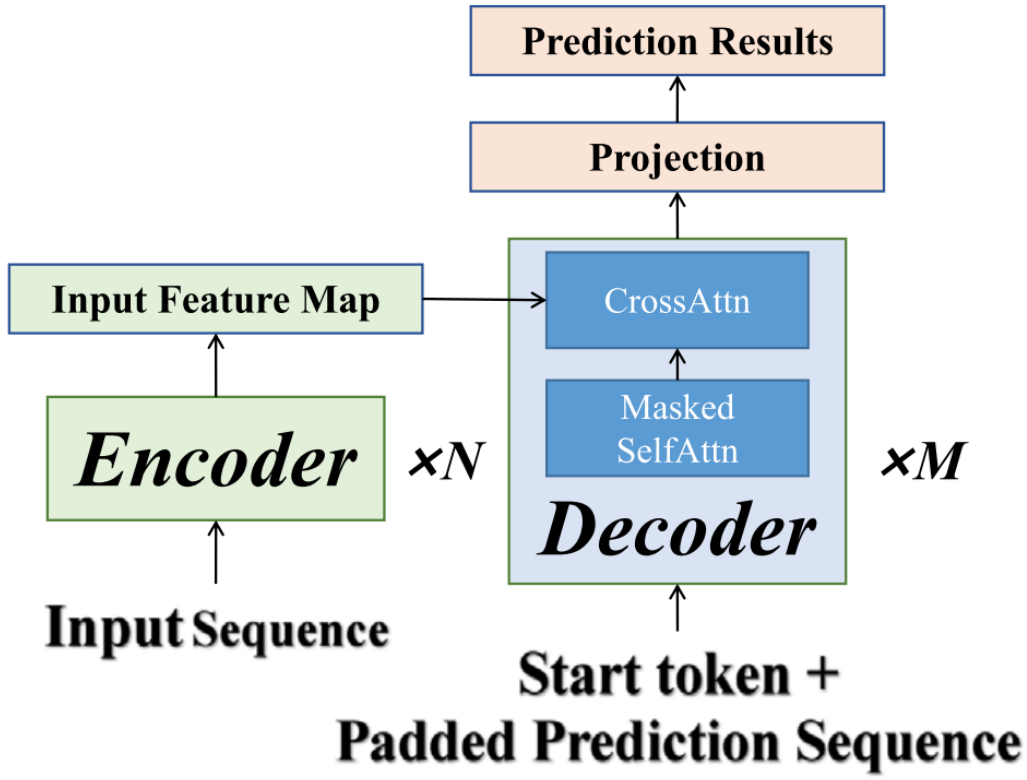
- **Decoder Inputs 的初始化方法存在问题**
  - Start token：Informer



01 问题描述

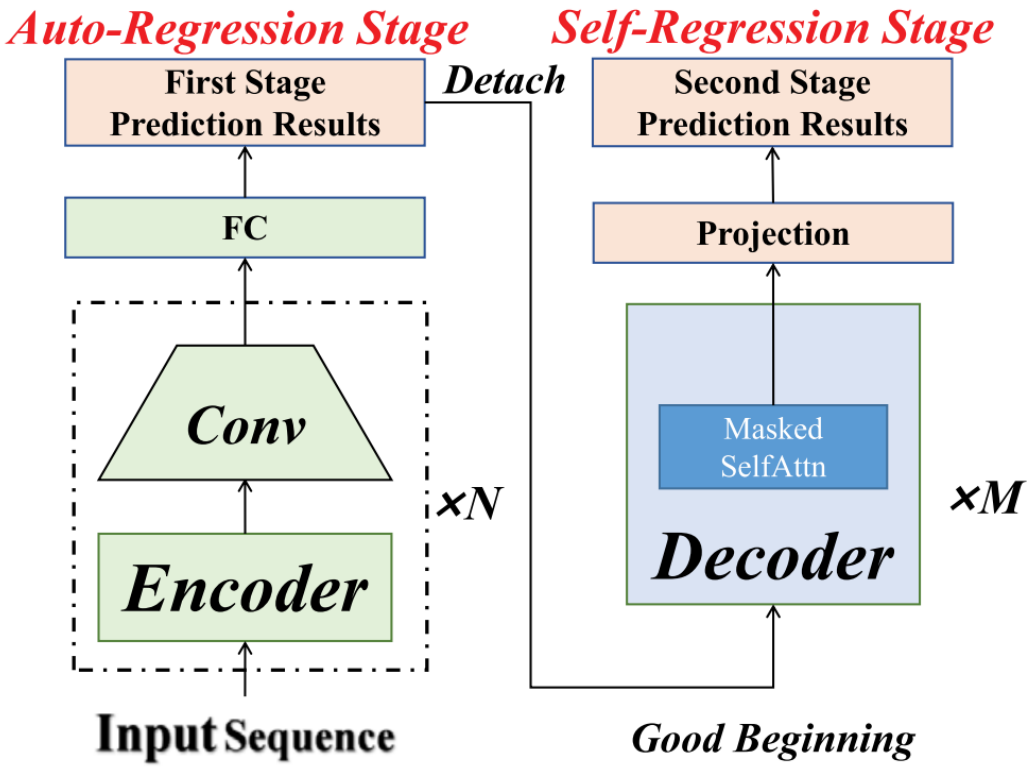
- **Decoder Inputs 的初始化方法存在问题**
  - Start token：Informer
  - 趋势分解方法：FEDformer





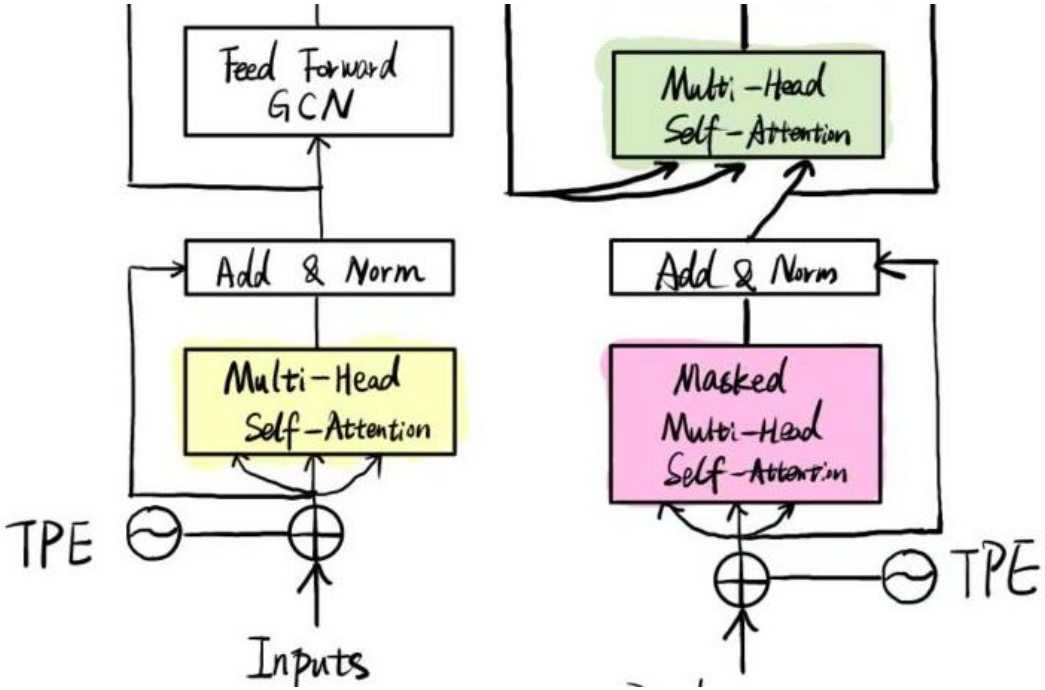
02 创新点

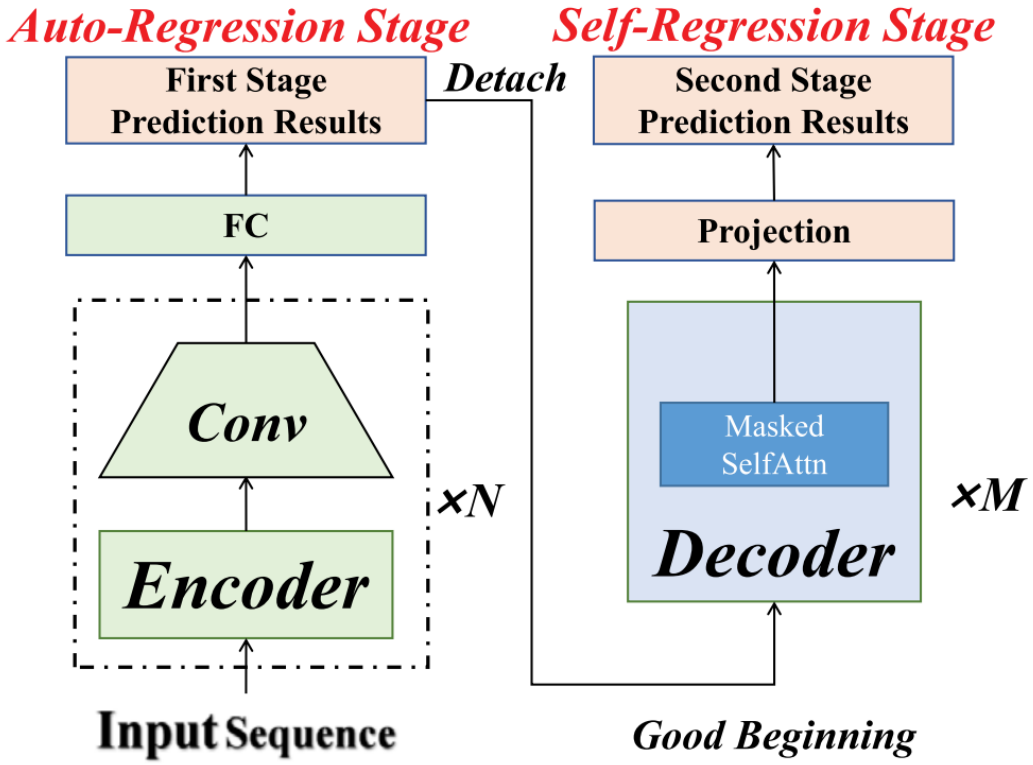
- **Two-stage** transformer framework：
  - 第一阶段：捕获输入序列的特征，使用FC层预测序列
  - 第二阶段：只关注预测序列内部的关系
- **GBT：Good Beginning Transformer**
  - 第一阶段的输出作为第二阶段的输入
- **Error Score Modification：**误差评分修改模块
  - 增强第二阶段的预测能力



02 创新点

- **Auto-Regression & Self-Regression**：
  - Auto-Regression：输入序列预测元素的推理过程
  - Self-Regression：预测元素本身的推理过程，从自身推导出预测元素

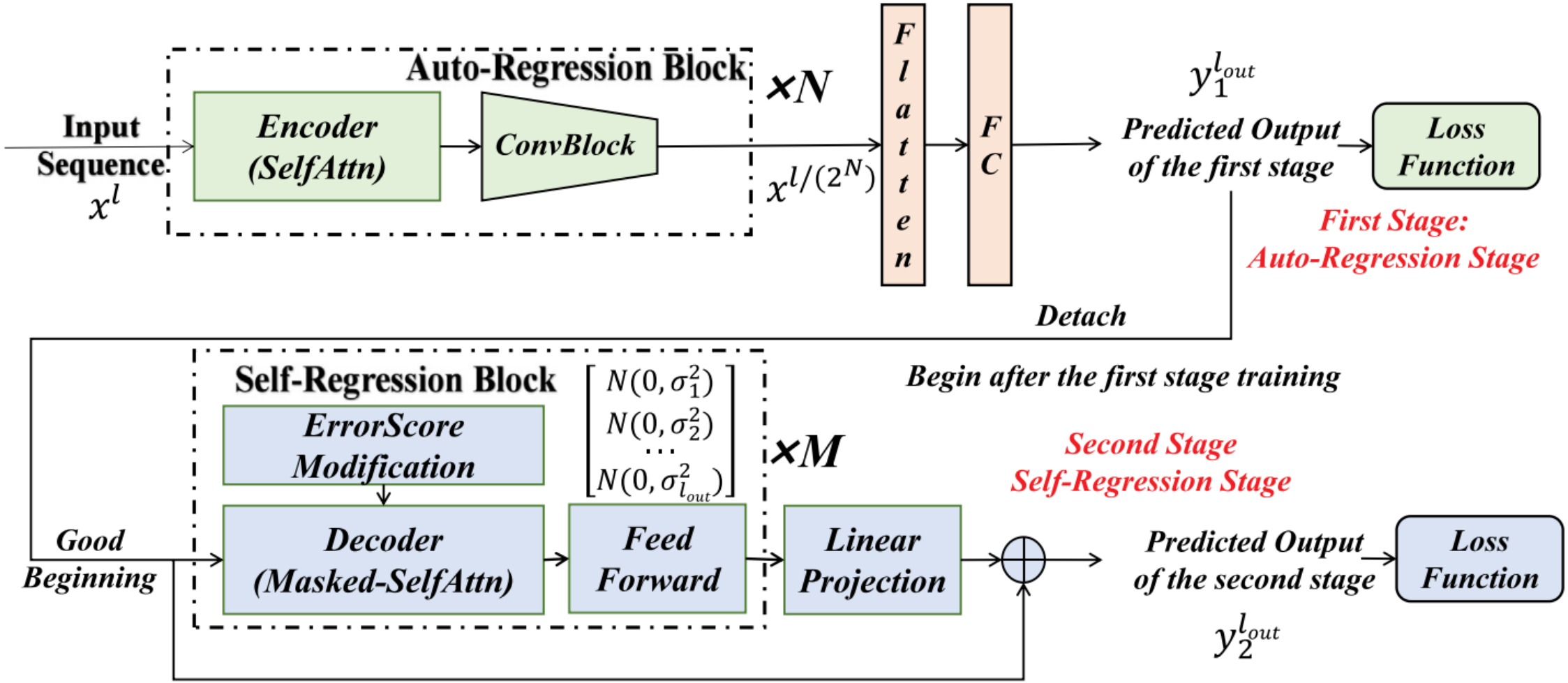




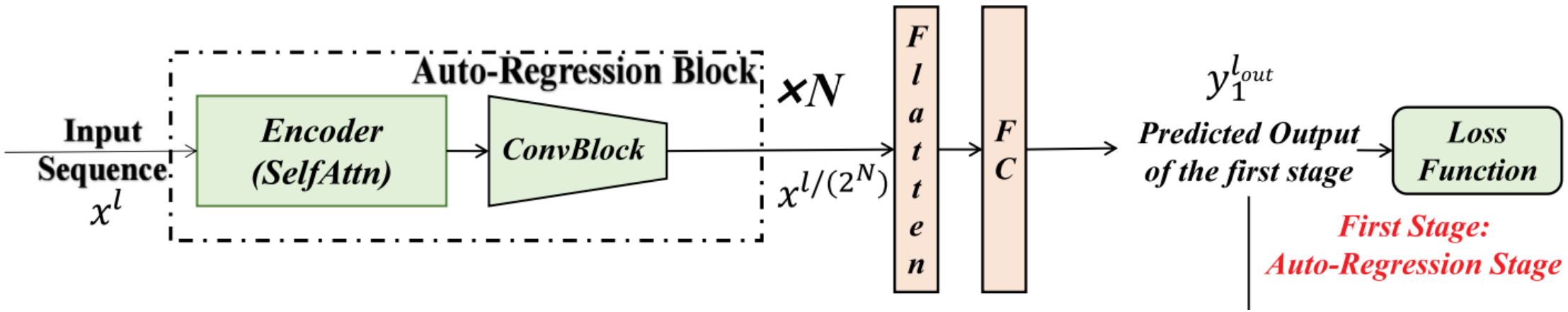
03 相关工作

- **两阶段预测方法**：（基于表征学习的自监督时间序列预测方法）
  - 第一阶段：通过对比学习获得输入序列的通用特征映射，
  - 第二阶段：通过第一阶段和回归器输出的表征获得预测结果。
- **GBT与传统的两阶段模型有以下不同：**
  - 两个阶段的GBT输出都是预测结果，而不是表征。
  - 在GBT中使用两阶段架构的目的是解决非平稳性问题，而不是增强网络的特征提取能力。
  - GBT是建立在Transformer之上的，而这些自监督方法主要是基于CNN/MLP。

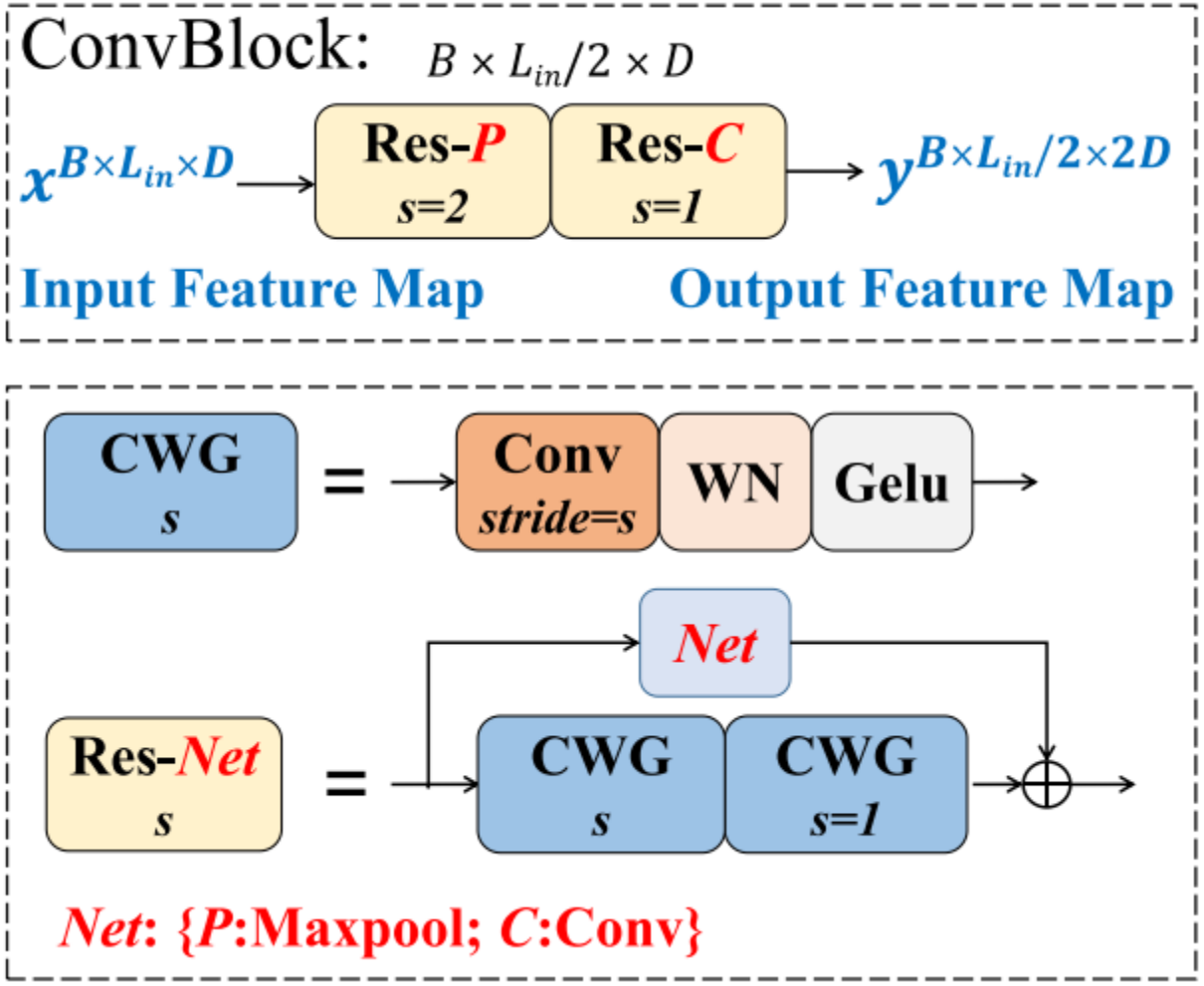
04 GBT: Overview



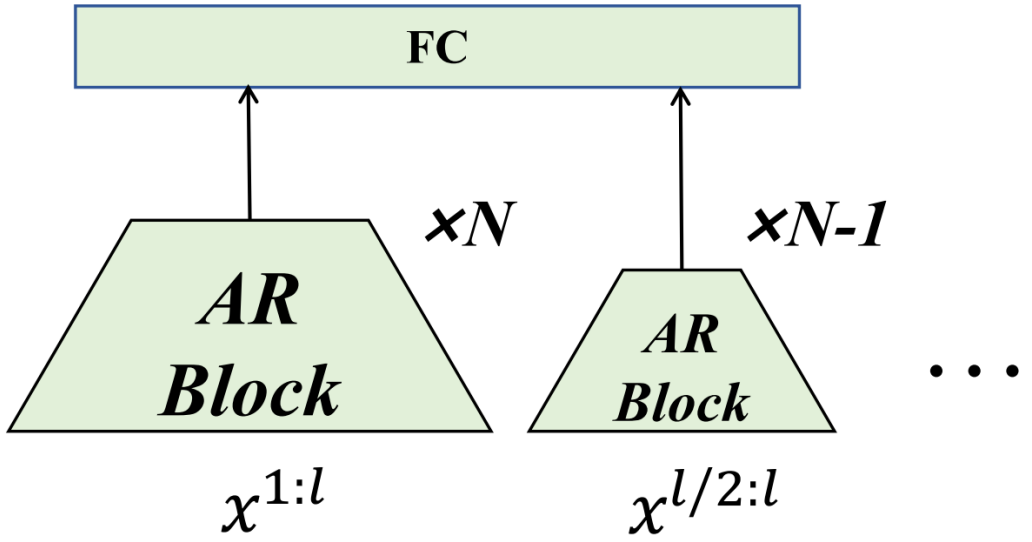
04 GBT：Auto-regression stage



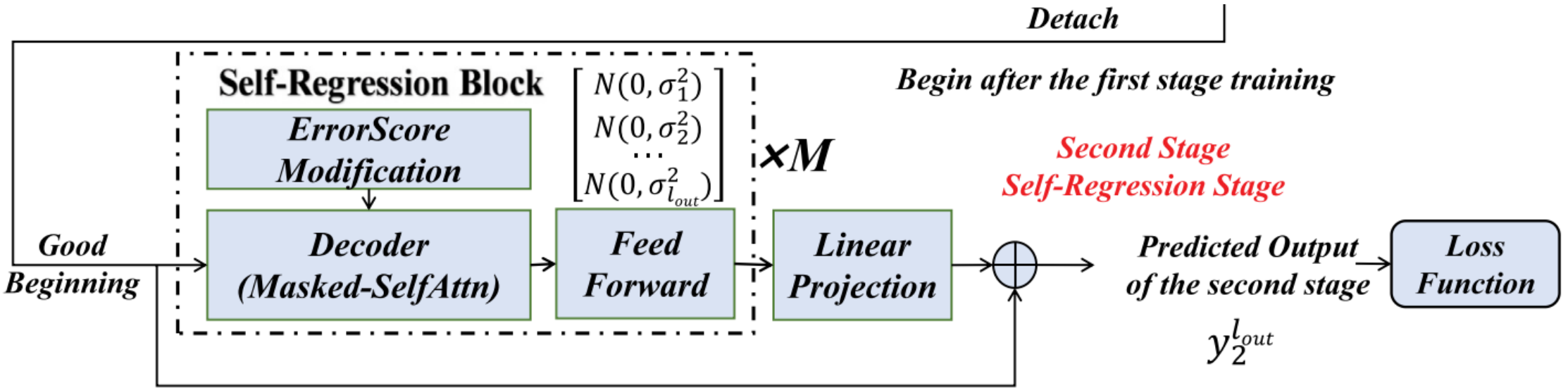
04 GBT：Auto-regression stage



Res-P：将序列长度缩短一半
Res-C：将隐藏维数加倍。



04 GBT：Self-regression stage

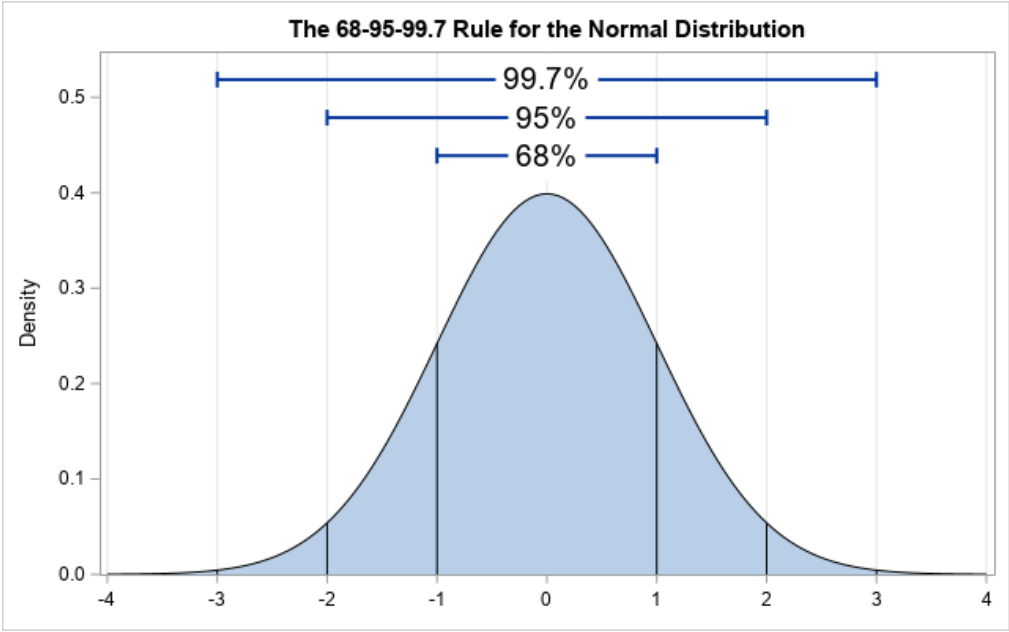


- 第二阶段的网络不参与第一阶段的训练阶段，第一阶段的参数将在第二阶段的训练阶段固定。
- ESM模块：合理地利用第一阶段的预测结果

04 GBT：Self-regression stage

- 引导第二阶段将更多的分数分配给更可能接近真实值的**早期元素**

$$ESM(Q,K) = QK^{\top} + \begin{bmatrix} N(0,\sigma_1^2) \\ N(0,\sigma_2^2) \\ \cdots \\ N(0,\sigma_{l_{out}}^2) \end{bmatrix}$$



05 实验1：对比实验

Methods	Metrics	WTH				Traffic				Exchange			
		96	192	336	720	96	192	336	720	96	192	336	720
GBT-Vanilla	MSE	***0.434***	***0.481***	***0.514***	***0.523***	***0.509***	***0.520***	***0.535***	***0.575***	0.110	***0.179***	0.358	***0.756***
	MAE	***0.466***	***0.506***	***0.527***	***0.532***	***0.282***	***0.293***	***0.307***	***0.317***	0.249	*0.312*	0.446	***0.655***
FEDformer	MSE	0.509	0.581	0.630	0.580	0.587	0.604	0.621	0.626	0.148	0.271	0.460	1.195
	MAE	0.513	0.557	0.636	0.586	*0.366*	0.373	0.383	0.382	0.278	0.380	0.500	0.841
Pyraformer	MSE	0.540	0.575	0.593	0.623	0.938	0.939	0.948	–	1.489	1.642	1.744	2.080
	MAE	0.546	0.567	0.578	0.599	0.490	0.488	0.488	–	1.018	1.075	1.107	1.197
ETSformer	MSE	0.538	0.615	0.655	0.719	0.607	0.621	0.622	0.632	***0.085***	*0.182*	***0.348***	1.025
	MAE	0.521	0.566	0.589	0.624	0.392	0.399	0.396	0.396	***0.204***	***0.303***	***0.428***	0.774
AirFormer	MSE	0.504	0.562	0.580	0.619	0.849	0.856	0.866	1.030	0.938	1.000	1.164	1.720
	MAE	0.520	0.560	0.570	0.595	0.476	0.478	0.480	0.613	0.840	0.870	0.932	1.115
SCINet	MSE	0.489	0.526	0.572	0.617	0.581	0.595	–	–	0.221	0.323	0.661	2.691
	MAE	0.495	0.524	0.562	0.586	0.423	0.429	–	–	0.365	0.442	0.564	1.320
TS2Vec	MSE	*0.450*	*0.505*	*0.532*	*0.566*	0.941	–	–	–	0.184	0.373	0.666	2.941
	MAE	*0.472*	*0.515*	*0.533*	*0.557*	0.550	–	–	–	0.315	0.452	0.612	1.313
DLinear	MSE	0.539	0.592	0.610	0.653	0.725	0.665	0.674	0.716	0.219	0.350	0.563	1.076
	MAE	0.522	0.557	0.571	0.600	0.460	0.438	0.441	0.457	0.387	0.481	0.606	0.799
N-HiTS	MSE	0.488	0.539	0.565	0.628	*0.561*	*0.543*	*0.554*	*0.616*	*0.099*	0.297	0.576	1.288
	MAE	0.496	0.536	0.555	0.596	0.369	*0.352*	*0.357*	*0.377*	*0.225*	0.390	0.550	0.847
FiLM	MSE	0.515	0.585	0.619	0.688	1.409	1.412	1.428	1.451	0.141	0.216	*0.351*	*0.938*
	MAE	0.497	0.540	0.563	0.602	0.799	0.802	0.806	0.809	0.272	0.342	*0.434*	*0.736*

05 实验2：消融实验

Methods	Metrics	$ETTh_1$ (Univariate)				$ETTh_1$ (Multivariate)			
		96	192	336	720	96	192	336	720
GBT-first	MSE	*0.060*	*0.077*	0.096	*0.128*	0.406	0.472	0.515	0.555
	MAE	*0.186*	0.210	0.240	*0.282*	0.419	0.461	0.471	*0.512*
GBT-second	MSE	***0.051***	***0.074***	***0.080***	***0.119***	0.398	*0.448*	0.497	***0.538***
	MAE	***0.173***	***0.206***	***0.221***	***0.276***	*0.418*	*0.442*	0.470	***0.505***
GBT-wo-simul	MSE	0.092	0.167	0.189	0.284	0.412	0.449	0.500	*0.543*
	MAE	0.237	0.331	0.342	0.455	0.431	0.448	0.476	0.521
GBT-wo-ESM	MSE	0.061	0.076	*0.083*	0.134	0.407	0.467	0.483	0.564
	MAE	0.186	*0.209*	*0.223*	0.304	0.420	0.458	*0.459*	0.512
GBT-wo-CB	MSE	0.085	0.079	0.149	0.246	***0.381***	***0.431***	***0.474***	0.614
	MAE	0.223	0.213	0.317	0.420	***0.401***	***0.422***	***0.452***	0.523
GBT-wo-Pyra	MSE	0.064	0.078	0.098	0.156	0.416	0.478	*0.488*	0.563
	MAE	0.193	0.212	0.243	0.326	0.431	0.468	0.468	0.516

1) GBT-first: GBT只保留第一阶段;
2) GBT-second: GBT-Vanilla;
3) GBT-simul: GBTVanilla，两阶段同时训练。
4) GBT-woESM：无ESM的GBT;
5) GBT-wo-CB：第一阶段用前馈层代替ConvBlock的GBT;
6) GBT- wo - pyra:没有金字塔网络的GBT。

05 实验3：验证cross-attention和start token是冗余的

Methods	Metrics	$ETTh_1$ (Univariate)				$ETTh_1$ (Multivariate)			
		96	192	336	720	96	192	336	720
GBT-second	MSE	***0.051***	0.074	*0.080*	***0.119***	***0.398***	***0.448***	*0.497*	***0.538***
	MAE	***0.173***	*0.206*	*0.221*	***0.276***	***0.418***	***0.442***	0.470	***0.505***
GBT-w-cross	MSE	*0.061*	***0.073***	0.095	0.235	0.403	*0.449*	***0.465***	0.584
	MAE	*0.187*	***0.204***	0.238	0.408	0.420	*0.442*	***0.442***	*0.515*
GBT-w-st	MSE	0.063	*0.074*	***0.076***	*0.166*	0.404	0.463	0.498	*0.582*
	MAE	0.192	0.207	***0.221***	*0.331*	0.420	0.460	*0.465*	0.534

05 实验4：验证通用性和适应性

Methods	Metrics	WTH				WTH			
		96	192	336	720	96	192	336	720
GBT-Vanilla	MSE	*0.188*	***0.221***	***0.239***	***0.218***	***0.434***	***0.481***	***0.514***	***0.523***
	MAE	*0.318*	***0.348***	***0.372***	***0.349***	***0.466***	***0.506***	***0.527***	***0.532***

Methods	Metrics	WTH (Univariate)				WTH (Multivariate)			
		96	192	336	720	96	192	336	720
GBT + FEDformer	MSE	***0.171***	***0.200***	***0.204***	***0.202***	***0.449***	***0.489***	***0.504***	***0.508***
	MAE	***0.303***	***0.331***	***0.334***	***0.332***	***0.473***	***0.500***	***0.510***	***0.515***
GBT + ETSformer	MSE	***0.202***	***0.227***	***0.234***	***0.231***	***0.444***	***0.494***	***0.506***	***0.532***
	MAE	***0.329***	***0.359***	***0.350***	***0.369***	***0.474***	***0.515***	***0.520***	***0.544***
GBT + SCINet	MSE	***0.200***	***0.230***	***0.232***	***0.212***	***0.436***	***0.484***	***0.507***	***0.509***
	MAE	***0.325***	***0.351***	***0.352***	***0.339***	***0.471***	***0.504***	***0.522***	***0.519***
GBT + DLinear	MSE	***0.196***	***0.235***	***0.237***	***0.225***	***0.497***	***0.547***	***0.577***	***0.627***
	MAE	***0.324***	***0.358***	***0.359***	***0.352***	***0.509***	***0.543***	***0.568***	***0.595***

05

实验5：验证处理非平稳序列的有效性

Methods	Metrics	$ETTh_1$ (Multivariate)				$ETTm_1$ (Multivariate)			
		168	336	720	960	48	96	288	672
Informer	MSE	1.138	1.278	1.357	1.470	0.499	0.605	0.906	0.943
	MAE	0.853	0.909	0.945	0.990	0.486	0.554	0.738	0.760
Informer + RevIN	MSE	*0.655*	*1.058*	*0.926*	*0.902*	*0.390*	*0.405*	*0.563*	*0.663*
	MAE	*0.561*	*0.758*	*0.717*	*0.715*	*0.391*	*0.411*	*0.502*	*0.550*
Informer + GBT	MSE	***0.443***	***0.511***	***0.555***	***0.596***	***0.268***	***0.308***	***0.363***	***0.429***
	MAE	***0.440***	***0.473***	***0.512***	***0.528***	***0.337***	***0.366***	***0.397***	***0.442***

谢谢观看

MANY THANKS !

23.12.12