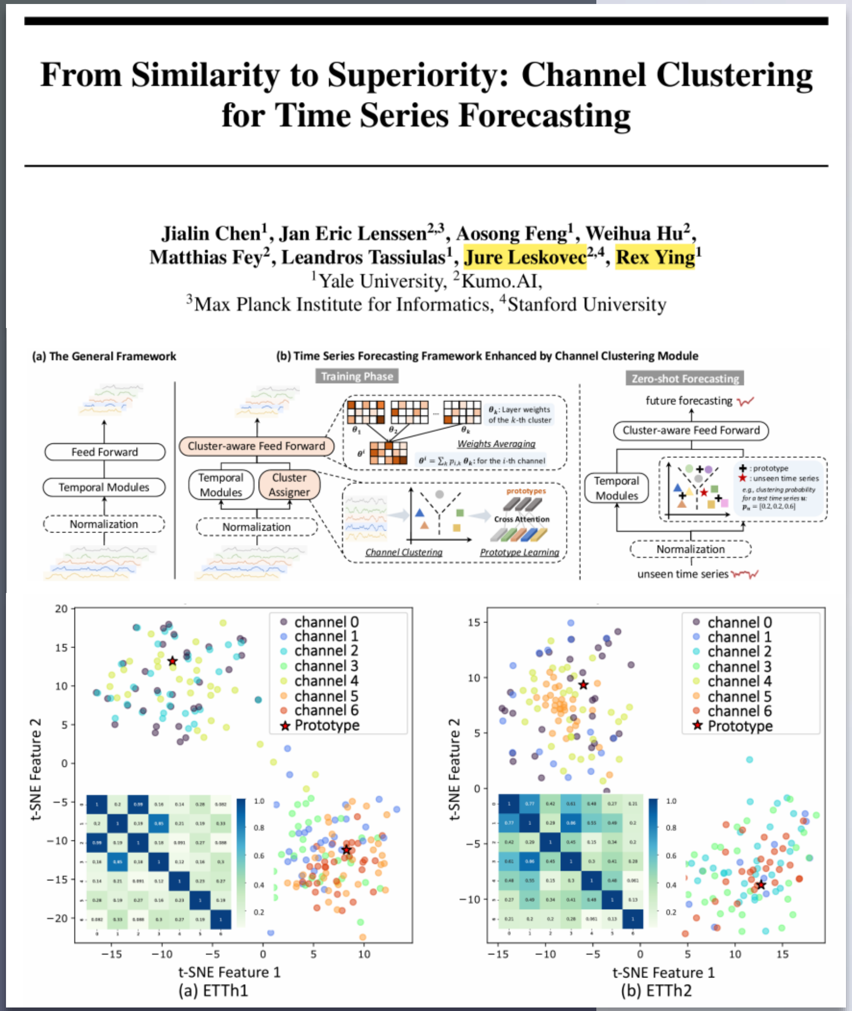

From Similarity to Superiority: Channel Clustering for Time Series Forecasting

Jialin Chen[1], Jan Eric Lenssen[2,3], Aosong Feng[1], Weihua Hu[2],
Matthias Fey[2], Leandros Tassiulas[1], Jure Leskovec[2,4], Rex Ying[1]
[1]Yale University, [2]Kumo.AI,
[3]Max Planck Institute for Informatics, [4]Stanford University

时间序列预测

通道独立

通道混合

通道聚类

NIPS 2024

2024 NIPS CCF-A

From Similarity to Superiority

Channel **C**lustering for Time Series Forecasting

24.11.28

Presented by Yyyq

01 问题描述

- **通道独立 CI：**每个通道单独建模 **泛化性和鲁棒性有限**

$$f^{(i)}:\mathbb{R}^T \to \mathbb{R}^H \text{ for } i=1,\cdots,C,$$

- **通道混合 CD：**所有通道整体建模 **导致过平滑**

$$f:\mathbb{R}^{T\times C} \to \mathbb{R}^{H\times C}$$

需要：平衡单个通道处理以提高预测性能，且不会忽略通道之间的基本交互。

02 创新点

- **通道聚类模块 CCM**：结合CD和CI的优点，即插即用
  - 分别提高 CI和CD的预测性能
  - 实现zero-shot预测
  - 提高时序模型的**可解释性**

03 动机

- **random shuffling 随机洗牌实验：**TSMixer和TimesNet、DLinear和PatchTST

$\Delta\mathcal{L}(\%)$ 均方误差（MSE）损失

PCC 皮尔逊相关系数

$$\begin{cases} \textsc{Sim}(X_i, X_j) = \exp\left(\frac{-\|X_i - X_j\|^2}{2\sigma^2}\right) \\ \Delta\mathcal{L}_{ij} := |\Delta\mathcal{L}_i - \Delta\mathcal{L}_j| \end{cases}$$

通道间相似度

通道间性能损失差

Base Model Channel Strategy		TSMixer *CD*	DLinear *CI*	PatchTST *CI*	TimesNet *CD*
ETTh1	$\Delta\mathcal{L}(\%)$	2.67	1.10	11.30	18.90
ETTm1	$\Delta\mathcal{L}(\%)$	4.41	5.55	6.83	14.98
Exchange	$\Delta\mathcal{L}(\%)$	16.43	19.34	27.98	24.57

(1) 现有的预测方法严重依赖于信道身份信息。

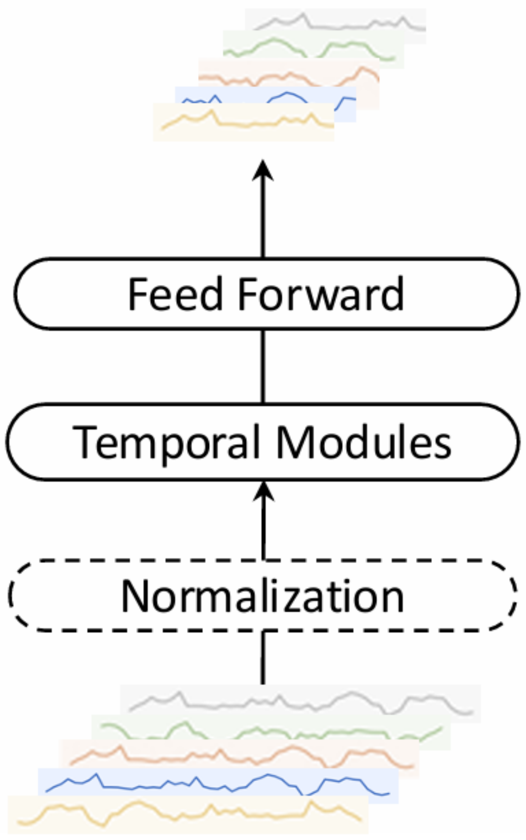
(2)这种依赖与通道相似性呈负相关

提供聚类身份而非通道身份

04

算法实现：整体结构

(a) The General Framework

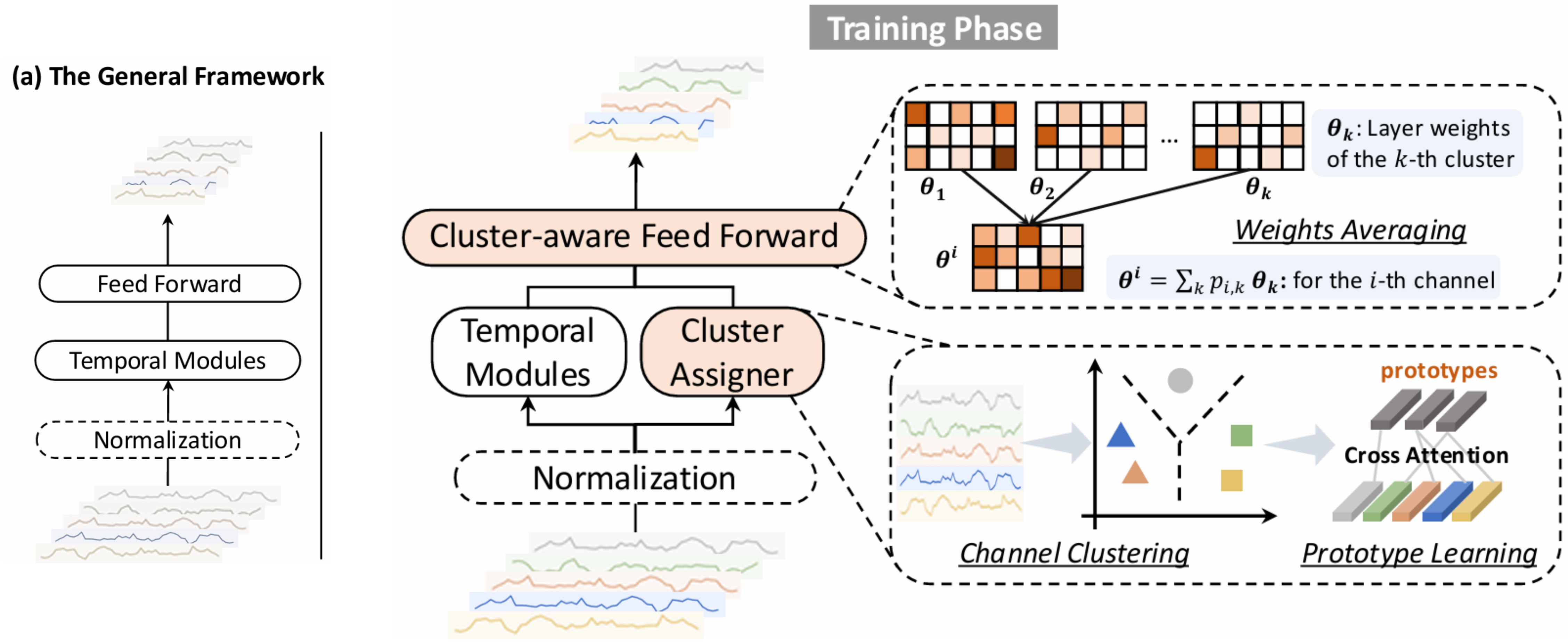


前馈层预测未来值

时序模块
Transformer/CNN/...

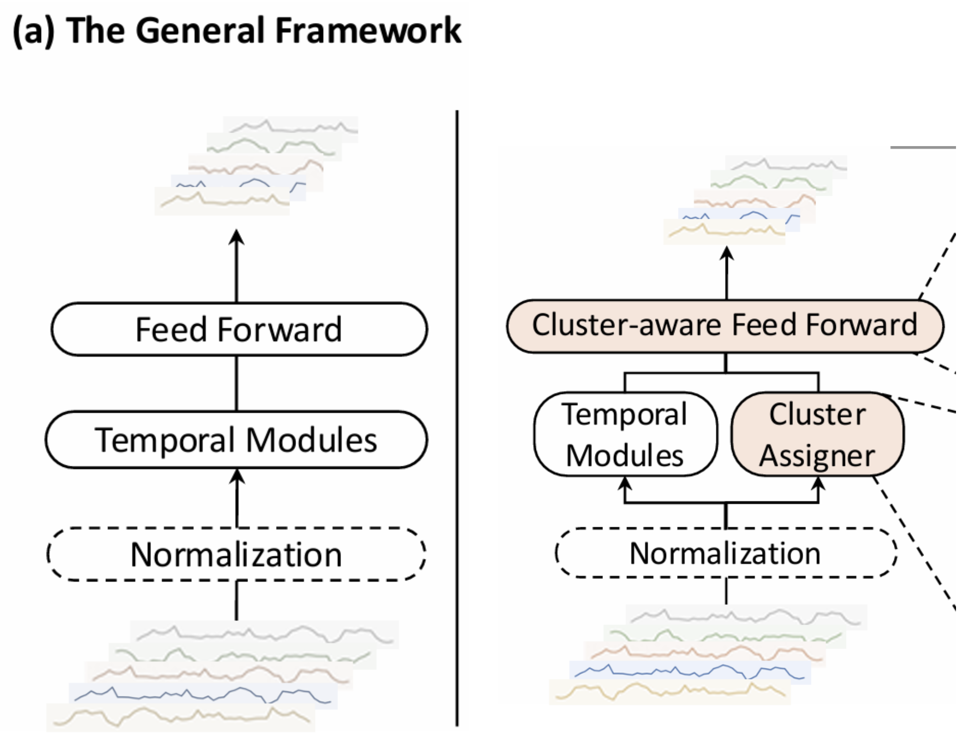
（可选）归一化层

04 算法实现：整体结构

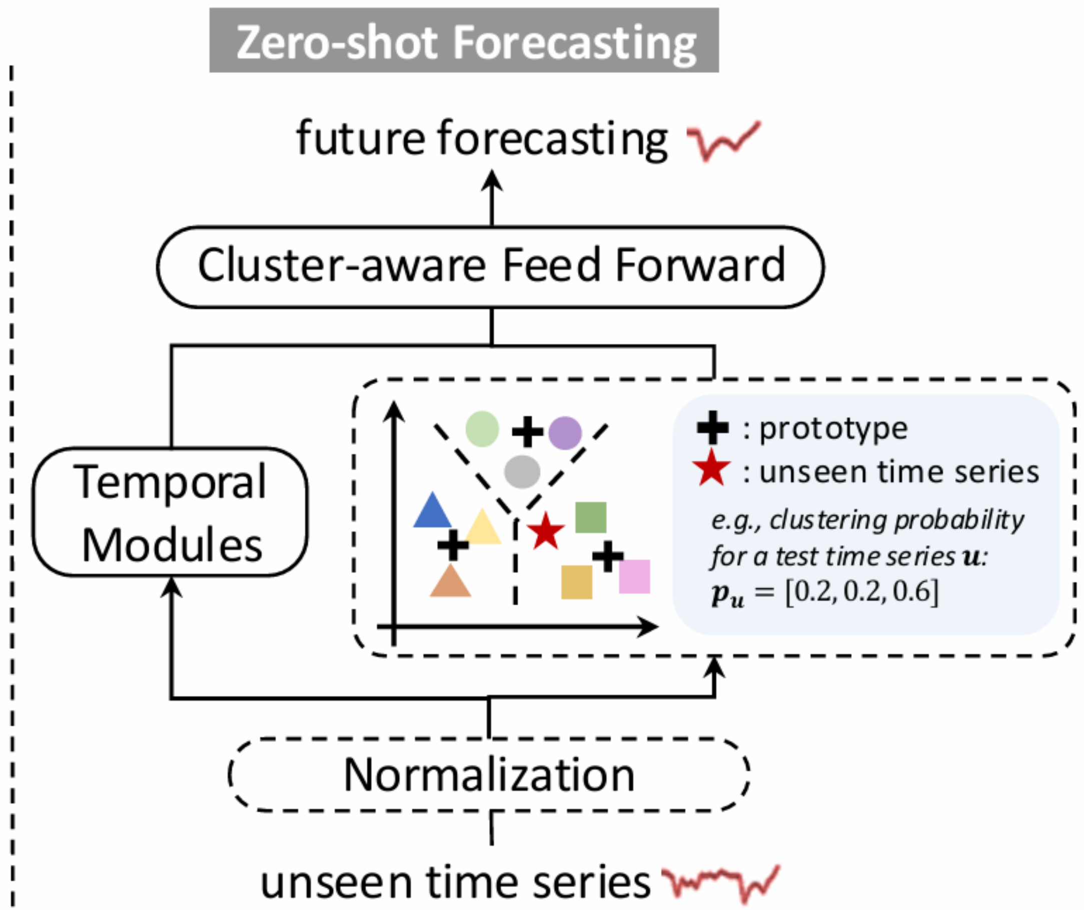


(a) The General Framework
Feed Forward
Temporal Modules
Normalization
Training Phase
Cluster-aware Feed Forward
Temporal Modules
Cluster Assigner
Normalization
θ_k: Layer weights of the k-th cluster
θ_1
θ_2
θ_k
θ^i
Weights Averaging
$\theta^i = \sum_k p_{i,k}\, \theta_k$: for the i-th channel
prototypes
Cross Attention
Channel Clustering
Prototype Learning

04 **算法实现：**整体结构



(a) The General Framework
Feed Forward
Temporal Modules
Normalization
Cluster-aware Feed Forward
Temporal Modules
Cluster Assigner
Normalization



Zero-shot Forecasting
future forecasting
Cluster-aware Feed Forward
Temporal Modules
: prototype
: unseen time series
e.g., clustering probability for a test time series u:
$p_u = [0.2, 0.2, 0.6]$
Normalization
unseen time series

04

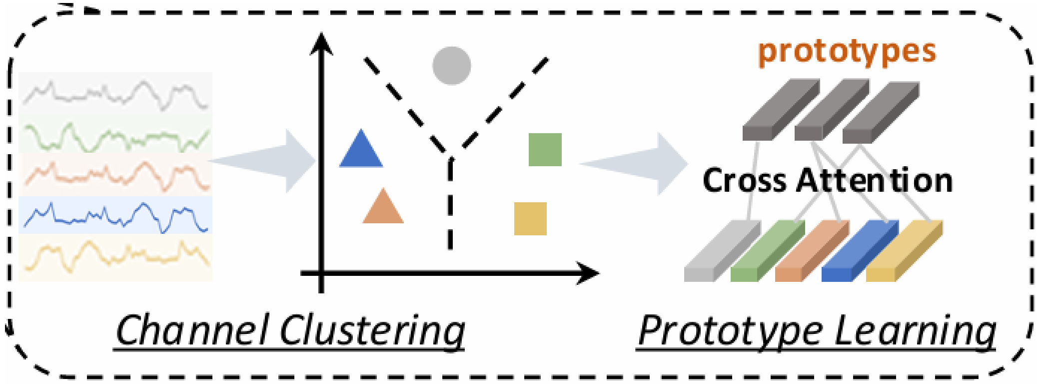
算法实现：CCM （Channel Clustering Module)

初始化聚类中心

分配数据点

更新质心

迭代更新



04 算法实现：CCM （Channel Clustering Module)

$X \in \mathbb{R}^{T \times C} \xrightarrow{MLP} \boxed{H = [h_1, \cdots, h_C] \in \mathbb{R}^{C \times d}}$ channel hidden emb

C 个 Channel

初始化聚类中心

$\boxed{C = [c_1, \cdots, c_K] \in \mathbb{R}^{K \times d}}$ cluster emb

K 个 Cluster

分配数据点

聚类概率 $p_{i,k} = \text{Normalize}(\frac{c_k^\top h_i}{\|c_k\| \|h_i\|}) \in [0,1].$ 概率分布

归一化内积

聚类成员矩阵 $M \in \mathbb{R}^{C \times k}$

$\begin{cases} C \text{个 channel} \\ K \text{个 cluster.} \end{cases}$

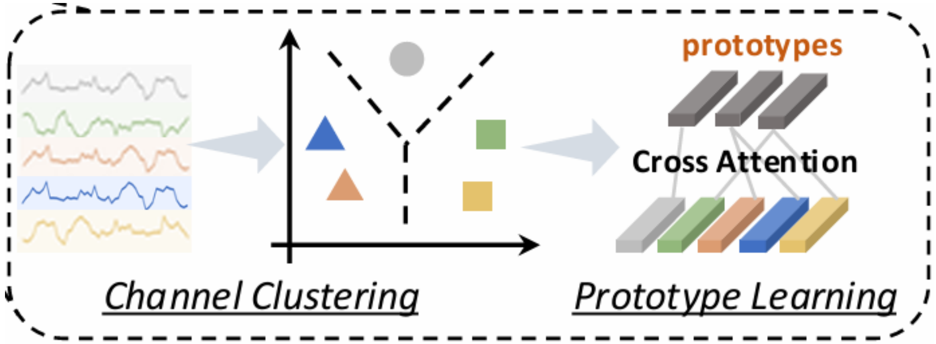
近似二值矩阵 (0,1)

$M_{i,k} = \text{Bernoulli}(p_{i,k})$

04 算法实现：CCM （Channel Clustering Module)

更新质心（原型学习）

$$\begin{cases} \mathbf{C} = [c_1, \cdots, c_K] \in \mathbb{R}^{K\times d} \\ \mathbf{H} = [h_1, \cdots, h_C] \in \mathbb{R}^{C\times d} \end{cases}$$



$$\widehat{\mathbf{C}} = \text{Normalize}\left(\exp\left(\frac{(W_Q\mathbf{C})(W_K\mathbf{H})^\top}{\sqrt{d}}\right) \odot \mathbf{M}^\top\right) W_V\mathbf{H},$$

$$\left(\frac{CH^T}{\sqrt{d}} \odot M^T\right) H$$

$\begin{cases} \text{query} : C \\ \text{Key} : H \end{cases}$, value : H.

04 算法实现：CCM （Channel Clustering Module)

迭代更新（聚类损失）

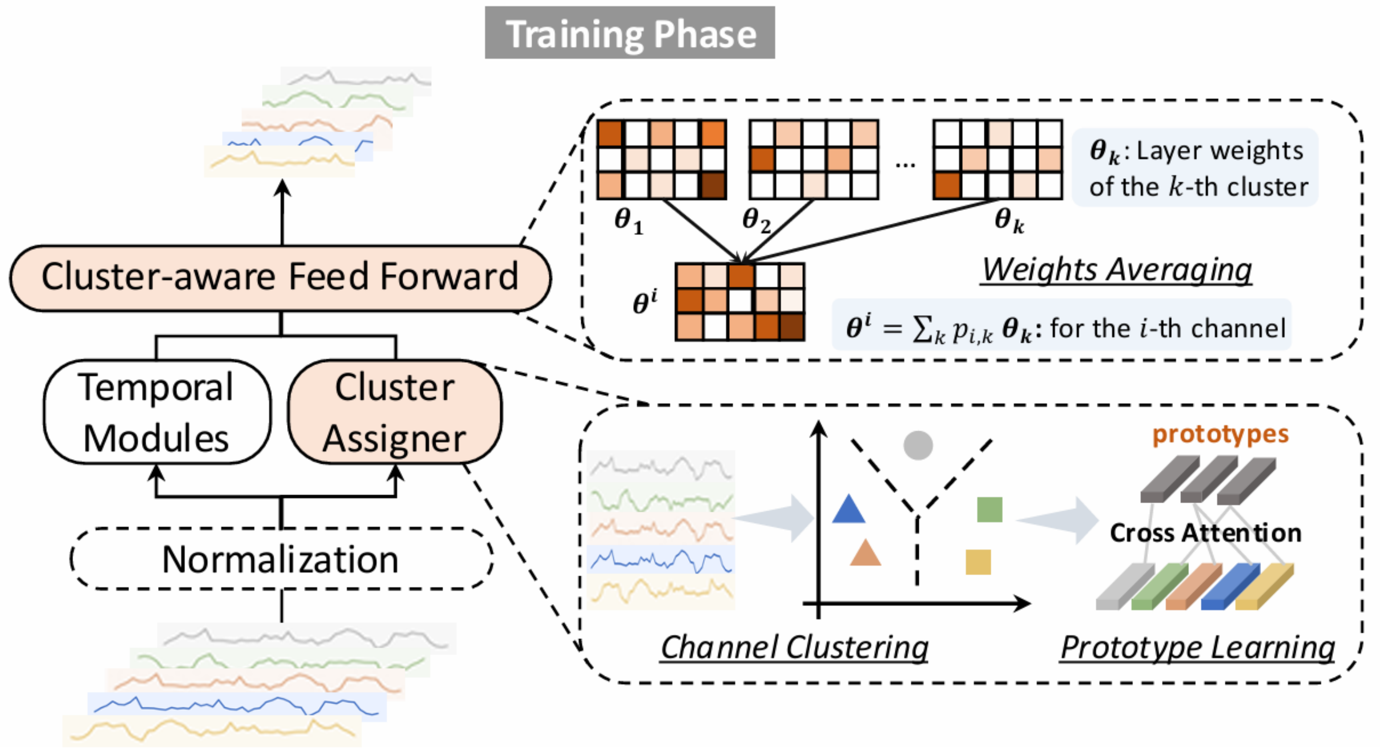
$$\begin{cases} \mathbf{M} \in \mathbb{R}^{C \times K} \\ \textsc{Sim}(X_i, X_j) = \exp(\frac{-\|X_i - X_j\|^2}{2\sigma^2}), \end{cases}$$

cluster loss

Tr 迹：方阵对角线元素之和.

$$\mathcal{L}_c = \underbrace{-Tr(M^T S M)}_{\text{最大化聚类内的相似性}} + \underbrace{Tr((I - MM^T)S)}_{\text{鼓励聚类之间的分离}}$$

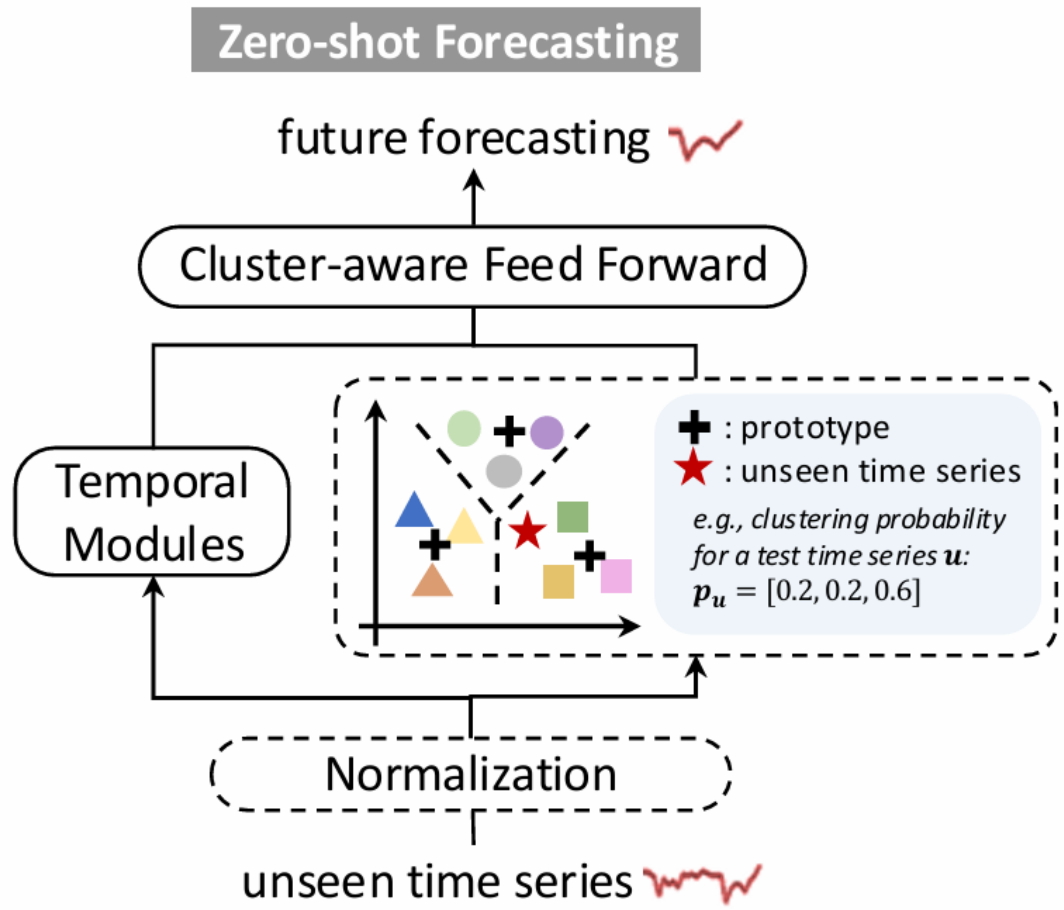
04 算法实现：CCM （Channel Clustering Module)



- 为每个聚类分配一个前馈网络
- 每个前馈网络为一个单独的线性层
- 最终预测：

  每个聚类的结果×相应的分类概率

04 算法实现：CCM （Channel Clustering Module)



零样本预测

- 取消交叉注意力（确定质心）

04 算法实现：CCM （Channel Clustering Module)

- **单变量适应性**
  - 从channels聚类扩展到samples聚类
- **复杂度分析**
  - 聚类计算：与聚类数量、通道数量和隐藏维度的乘积 成正比
  - 前馈计算：与聚类数量、通道数量和预测范围 成正比

05 实验：数据集 和 基准模型

Table 2: The statistics of datasets in long-term forecasting. Horizon is $\{96, 192, 336, 720\}$.

Dataset	Channels	Length	Frequency
ETTh1&ETTh2	7	17420	1 hour
ETTm1&ETTm2	7	69680	15 min
ILI	7	966	1 week
Exchange	8	7588	1 day
Weather	21	52696	10 min
Electricity	321	26304	1 hour
Traffic	862	17544	1 hour

Table 3: Dataset details of M4 and Stock in short-term forecasting.

Dataset	Length	Horizon
M4 Yearly	23000	6
M4 Quarterly	24000	8
M4 Monthly	48000	18
M4 Weekly	359	13
M4 Daily	4227	14
M4 Hourly	414	48
Stock (New)	10000	7/24

- TSMixer：CD，MLP-based
- Dlinear：CI，Linear-based
- PatchTST：CI，transformer-based
- TimesNet：CD，CNN-based

05

实验：长期预测效果

Dataset	ETTh1	ETTm1	ETTh2	ETTm2
Correlation r	0.1876	0.1717	0.3224	0.328

Exchange	**ILI**	**Weather**	**Electricity**	**Traffic**
0.3198	0.508	0.1169	0.5311	0.6325

Model		TSMixer		+ CCM		DLinear		+ CCM		PatchTST		+ CCM		TimesNet		+ CCM		IMP(%)
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
ETTh1	96	**0.361**	**0.392**	0.365	0.393	0.375	0.399	**0.371**	**0.393**	0.375	0.398	**0.371**	**0.396**	0.384	0.402	**0.380**	**0.400**	0.539
	192	0.404	**0.418**	**0.402**	**0.418**	0.405	0.416	**0.404**	**0.415**	0.415	0.425	**0.414**	**0.424**	0.436	0.429	**0.431**	**0.425**	0.442
	336	**0.422**	**0.430**	0.423	**0.430**	0.445	**0.440**	**0.438**	0.443	0.422	0.440	**0.417**	**0.429**	0.491	0.469	**0.485**	**0.461**	0.908
	720	0.463	0.472	**0.462**	**0.470**	0.489	**0.488**	**0.479**	0.497	0.449	**0.468**	**0.447**	0.469	0.521	0.500	**0.520**	**0.493**	0.333
ETTm1	96	0.285	0.339	**0.283**	**0.337**	0.299	**0.343**	**0.298**	**0.343**	0.294	0.351	**0.289**	**0.338**	0.338	0.375	**0.335**	**0.371**	1.123
	192	0.339	**0.365**	**0.336**	0.368	0.335	**0.365**	**0.334**	**0.365**	0.334	0.370	**0.333**	**0.363**	0.374	0.387	**0.373**	**0.383**	0.482
	336	0.361	0.406	**0.359**	**0.393**	0.370	0.386	**0.365**	**0.385**	0.373	0.397	**0.370**	**0.392**	**0.410**	**0.411**	0.412	0.416	0.716
	720	0.445	0.470	**0.424**	**0.421**	0.427	0.423	**0.424**	**0.417**	**0.416**	**0.420**	0.419	0.430	0.478	0.450	**0.477**	**0.448**	1.852
ETTh2	96	0.284	0.343	**0.278**	**0.338**	0.289	0.353	**0.285**	**0.348**	0.278	0.340	**0.274**	**0.336**	0.340	0.374	**0.336**	**0.371**	1.371
	192	0.339	**0.385**	**0.325**	0.393	0.384	0.418	**0.376**	**0.413**	0.341	0.382	**0.339**	**0.355**	0.402	0.414	**0.400**	**0.410**	1.806
	336	**0.361**	0.406	**0.361**	**0.399**	0.442	0.459	**0.438**	**0.455**	0.329	0.384	**0.327**	**0.383**	0.452	0.452	**0.449**	**0.445**	0.823
	720	0.445	0.470	**0.438**	**0.464**	0.601	0.549	**0.499**	**0.496**	0.381	0.424	**0.378**	**0.415**	0.462	0.468	**0.457**	**0.461**	4.370
ETTm2	96	0.171	**0.260**	**0.167**	**0.260**	0.167	0.260	**0.166**	**0.258**	0.174	0.261	**0.168**	**0.256**	**0.187**	**0.267**	0.189	0.270	0.860
	192	0.221	**0.296**	**0.220**	**0.296**	0.284	0.352	**0.243**	**0.323**	0.238	0.307	**0.231**	**0.300**	**0.249**	**0.309**	0.250	0.310	3.453
	336	**0.276**	**0.329**	0.277	0.330	0.369	0.427	**0.295**	**0.358**	0.293	0.346	**0.275**	**0.331**	0.321	0.351	**0.318**	**0.347**	6.012
	720	0.420	0.422	**0.369**	**0.391**	0.554	0.522	**0.451**	**0.456**	**0.373**	0.401	0.374	**0.400**	0.408	0.403	**0.394**	**0.391**	7.139
Exchange	96	0.089	0.209	**0.085**	**0.206**	0.088	0.215	**0.085**	**0.214**	0.094	0.216	**0.088**	**0.208**	0.107	0.234	**0.105**	**0.231**	2.880
	192	0.195	0.315	**0.177**	**0.300**	0.178	0.317	**0.171**	**0.306**	0.191	0.311	**0.185**	**0.309**	0.226	0.344	**0.224**	**0.340**	3.403
	336	0.343	0.421	**0.312**	**0.405**	0.371	0.462	**0.300**	**0.412**	0.343	0.427	**0.342**	**0.423**	0.367	0.448	**0.361**	**0.442**	5.875
	720	0.898	0.710	**0.847**	**0.697**	0.966	0.754	**0.811**	**0.683**	0.888	0.706	**0.813**	**0.673**	0.964	0.746	**0.957**	**0.739**	5.970
ILI	24	**1.914**	0.879	1.938	**0.874**	2.215	1.081	**1.935**	**0.935**	1.593	0.757	**1.561**	**0.750**	2.317	**0.934**	**2.139**	0.936	4.483
	36	1.808	0.858	**1.800**	**0.851**	2.142	0.977	**1.938**	**0.942**	1.768	0.794	**1.706**	**0.780**	1.972	0.920	**1.968**	**0.914**	2.561
	48	1.797	0.873	**1.796**	**0.867**	2.335	1.056	**2.221**	**1.030**	1.799	0.916	**1.774**	**0.892**	2.238	0.940	**2.229**	**0.937**	1.602
	60	1.859	0.895	**1.810**	**0.876**	2.479	**1.088**	**2.382**	1.096	1.850	0.943	**1.735**	**0.880**	**2.027**	**0.928**	2.041	0.930	2.491
Weather	96	0.149	0.198	**0.147**	**0.194**	0.192	0.250	**0.187**	**0.245**	0.149	0.198	**0.147**	**0.197**	0.172	0.220	**0.169**	**0.215**	1.729
	192	0.201	0.248	**0.192**	**0.242**	0.248	0.297	**0.240**	**0.285**	0.194	0.241	**0.191**	**0.238**	0.219	0.261	**0.215**	**0.257**	2.539
	336	0.264	0.291	**0.244**	**0.281**	0.284	0.335	**0.274**	**0.324**	**0.244**	**0.282**	0.245	0.285	0.280	0.306	**0.274**	**0.291**	2.924
	720	0.320	0.336	**0.318**	**0.334**	0.339	0.374	**0.320**	**0.357**	0.320	0.334	**0.316**	**0.333**	**0.365**	**0.359**	0.366	0.362	1.476
Electricity	96	0.142	0.237	**0.139**	**0.235**	0.153	**0.239**	**0.142**	0.247	0.138	0.233	**0.136**	**0.231**	0.168	0.272	**0.158**	**0.259**	2.480
	192	0.154	0.248	**0.147**	**0.246**	0.158	0.251	**0.152**	**0.248**	**0.153**	**0.247**	**0.153**	0.248	0.184	0.289	**0.172**	**0.262**	3.226
	336	0.163	0.264	**0.161**	**0.262**	0.170	0.269	**0.168**	**0.267**	0.170	0.263	**0.168**	**0.262**	0.198	0.300	**0.181**	**0.284**	2.423
	720	0.208	0.300	**0.204**	**0.299**	0.233	0.342	**0.230**	**0.338**	**0.206**	**0.296**	0.210	0.301	0.220	0.320	**0.205**	**0.309**	1.417
Traffic	96	0.376	0.264	**0.375**	**0.262**	**0.411**	0.284	**0.411**	**0.282**	0.360	0.249	**0.357**	**0.246**	0.593	0.321	**0.554**	**0.316**	1.488
	192	0.397	**0.277**	**0.340**	0.279	0.423	0.287	**0.422**	**0.286**	**0.379**	0.256	**0.379**	**0.254**	0.617	0.336	**0.562**	**0.331**	3.175
	336	0.413	0.290	**0.411**	**0.289**	0.438	0.299	**0.436**	**0.297**	0.401	0.270	**0.389**	**0.255**	0.629	**0.336**	**0.579**	0.341	2.120
	720	0.444	0.306	**0.441**	**0.302**	**0.467**	**0.316**	0.471	0.318	0.443	0.294	**0.430**	**0.281**	0.640	**0.350**	**0.587**	0.366	1.445

05

实验：短期预测效果

Dataset	M4 Monthly	M4 Daily	M4 Yearly	M4 Hourly	M4 Quarterly	M4 Weekly
Correlation r	0.62	0.646	0.712	0.55	0.671	0.653

Model		TSMixer	+ CCM	DLinear	+ CCM	PatchTST	+ CCM	TimesNet	+ CCM	IMP(%)
M4 (Yearly)	SMAPE	14.702	**14.676**	16.965	**14.337**	13.477	**13.304**	15.378	**14.426**	7.286
	MASE	**3.343**	3.370	4.283	**3.144**	3.019	**2.997**	3.554	**3.448**	9.589
	OWA	0.875	**0.873**	1.058	**0.834**	0.792	**0.781**	0.918	**0.802**	11.346
M4 (Quarterly)	SMAPE	11.187	**10.989**	12.145	**10.513**	10.380	**10.359**	10.465	**10.121**	6.165
	MASE	1.346	**1.332**	1.520	**1.243**	1.233	**1.224**	1.227	**1.183**	7.617
	OWA	0.998	**0.984**	1.106	**0.931**	0.921	**0.915**	0.923	**0.897**	6.681
M4 (Monthly)	SMAPE	13.433	**13.407**	13.514	**13.370**	12.959	**12.672**	13.513	**12.790**	2.203
	MASE	1.022	**1.019**	1.037	**1.005**	0.970	**0.941**	1.039	**0.942**	4.238
	OWA	0.946	**0.944**	0.956	**0.936**	0.905	**0.895**	0.957	**0.891**	3.067
M4 (Others)	SMAPE	**7.067**	7.178	6.709	**6.160**	4.952	**4.643**	6.913	**5.218**	10.377
	MASE	5.587	**5.302**	4.953	**4.713**	3.347	**3.128**	4.507	**3.892**	7.864
	OWA	1.642	**1.536**	1.487	**1.389**	1.049	**0.997**	1.438	**1.217**	9.472
M4 (Avg.)	SMAPE	12.867	**12.807**	13.639	**12.546**	12.059	**11.851**	12.880	**11.914**	5.327
	MASE	1.887	**1.864**	2.095	**1.740**	1.623	**1.587**	1.836	**1.603**	10.285
	OWA	0.957	**0.948**	1.051	**0.917**	0.869	**0.840**	0.955	**0.894**	6.693
Stock (Horizon 7)	MSE	0.939	**0.938**	0.992	**0.883**	0.896	**0.892**	0.930	**0.915**	3.288
	MAE	0.807	**0.806**	0.831	**0.774**	**0.771**	**0.771**	0.802	**0.793**	2.026
Stock (Horizon 24)	MSE	1.007	**0.991**	0.996	**0.917**	0.930	**0.880**	0.998	**0.937**	5.252
	MAE	0.829	**0.817**	0.832	**0.781**	0.789	**0.765**	0.830	**0.789**	3.889

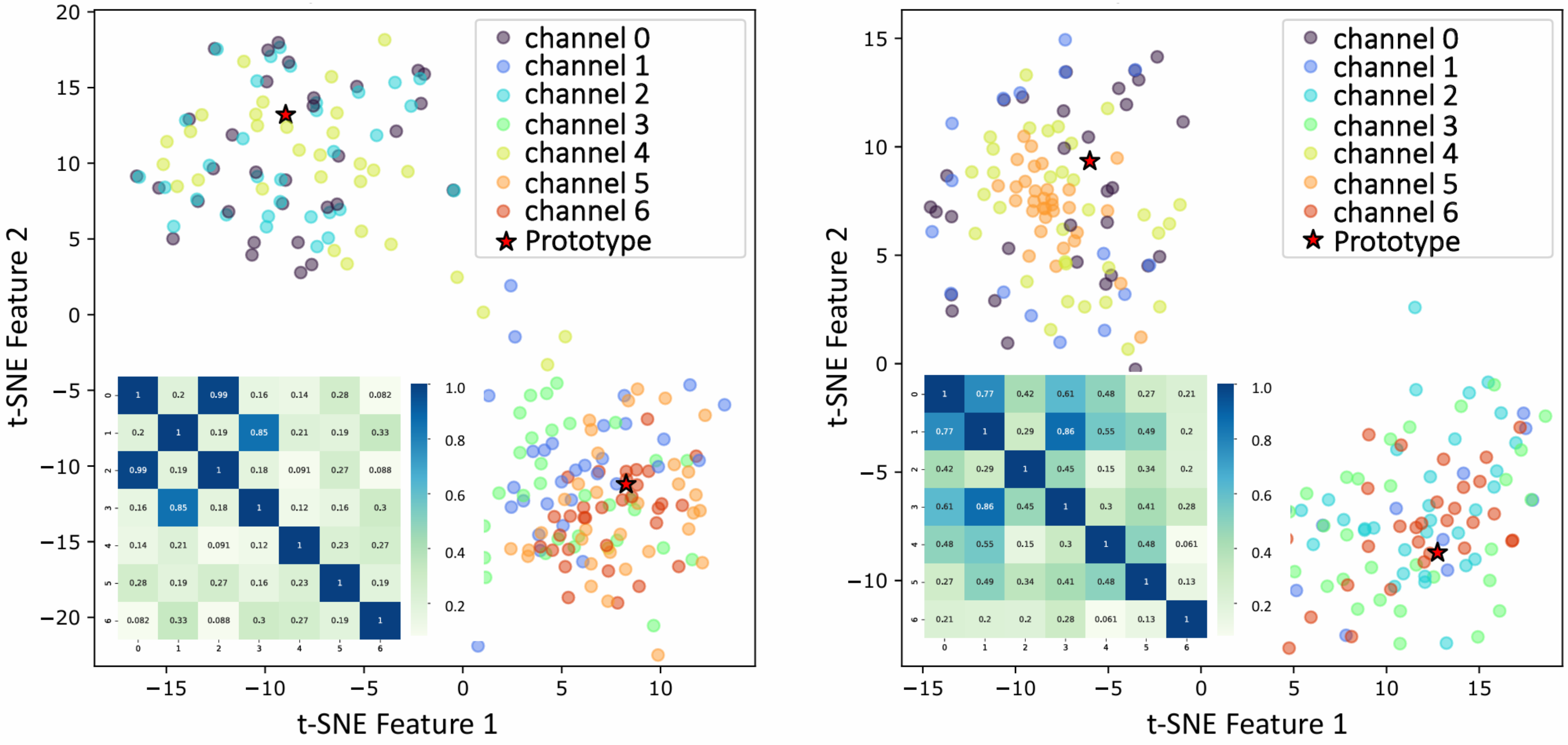
05

实验：零样本预测效果

Model Generalization Task		TSMixer		+ CCM		DLinear		+ CCM		PatchTST		+ CCM		TimesNet		+ CCM		IMP(%)
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
① ETTh1→ETTh2	96	0.288	0.357	0.283	0.353	0.308	0.371	0.283	0.349	0.313	0.362	0.292	0.346	0.391	0.412	0.388	0.410	3.661
	720	0.374	0.414	0.370	0.413	0.569	0.549	0.520	0.517	0.414	0.442	0.386	0.423	0.540	0.508	0.516	0.491	4.326
② ETTh1→ETTm1	96	0.763	0.677	0.710	0.652	0.726	0.658	0.681	0.634	0.729	0.667	0.698	0.647	0.887	0.718	0.827	0.700	4.626
	720	1.252	0.815	1.215	0.803	1.881	0.948	1.138	0.809	1.459	0.845	1.249	0.795	1.623	0.981	1.601	0.964	10.249
③ ETTh1→ETTm2	96	0.959	0.694	0.937	0.689	0.990	0.704	0.896	0.677	0.918	0.694	0.895	0.677	1.199	0.794	1.122	0.731	4.457
	720	1.765	0.982	1.758	0.980	2.091	1.061	1.681	0.954	1.925	1.014	1.718	0.966	2.204	1.031	1.874	1.012	7.824
④ ETTh2→ETTh1	96	0.466	0.462	0.455	0.456	0.462	0.450	0.427	0.432	0.620	0.563	0.509	0.495	0.869	0.624	0.752	0.590	8.016
	720	0.695	0.584	0.540	0.519	0.511	0.518	0.484	0.502	1.010	0.968	0.936	0.686	1.274	0.783	0.845	0.642	16.243
⑤ ETTh2→ETTm2	96	0.943	0.726	0.876	0.697	0.736	0.656	0.700	0.642	0.840	0.708	0.771	0.688	1.250	0.850	1.064	0.793	6.344
	720	1.472	0.872	1.464	0.866	1.813	0.938	1.253	0.844	1.832	1.052	1.532	0.863	1.861	1.016	1.671	0.967	11.439
⑥ ETTh2→ETTm1	96	1.254	0.771	1.073	0.714	1.147	0.746	0.894	0.669	0.997	0.721	0.789	0.629	1.049	0.791	0.804	0.657	16.016
	720	2.275	1.137	1.754	1.065	1.992	1.001	1.740	0.970	2.651	1.149	1.695	0.971	2.183	1.103	1.742	0.983	15.952

05

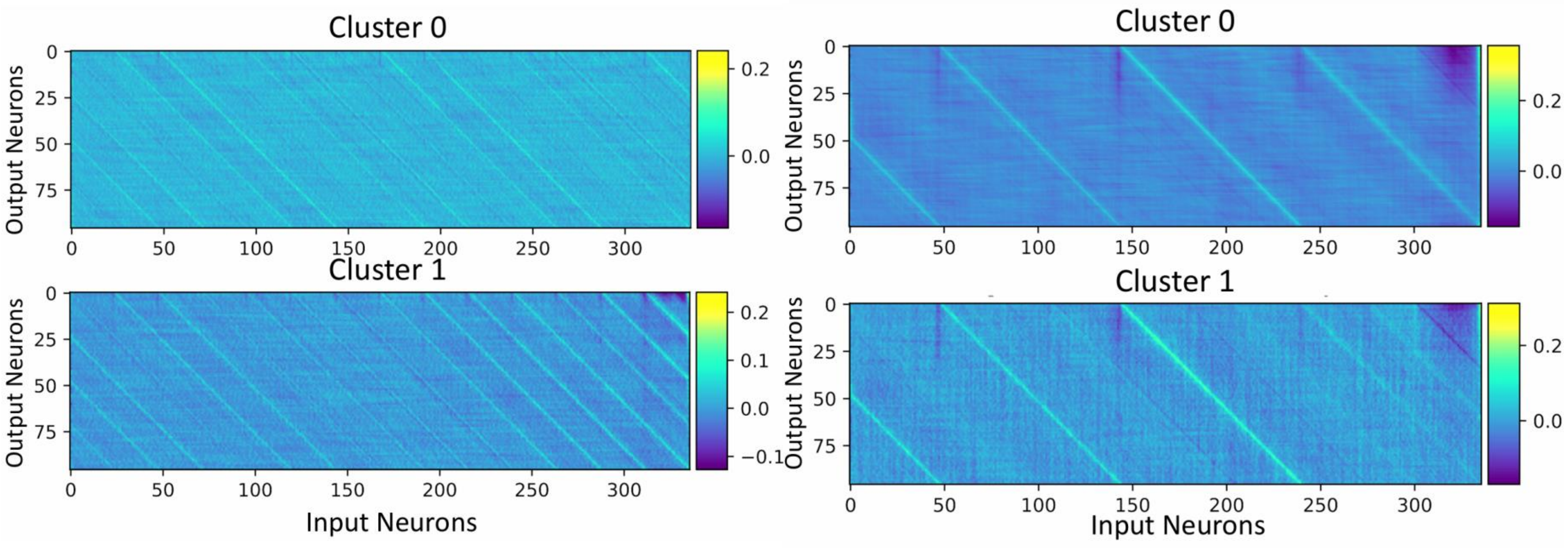
实验：可视化分析



(a) ETTh1 (b) ETTh2

05

实验：可视化分析

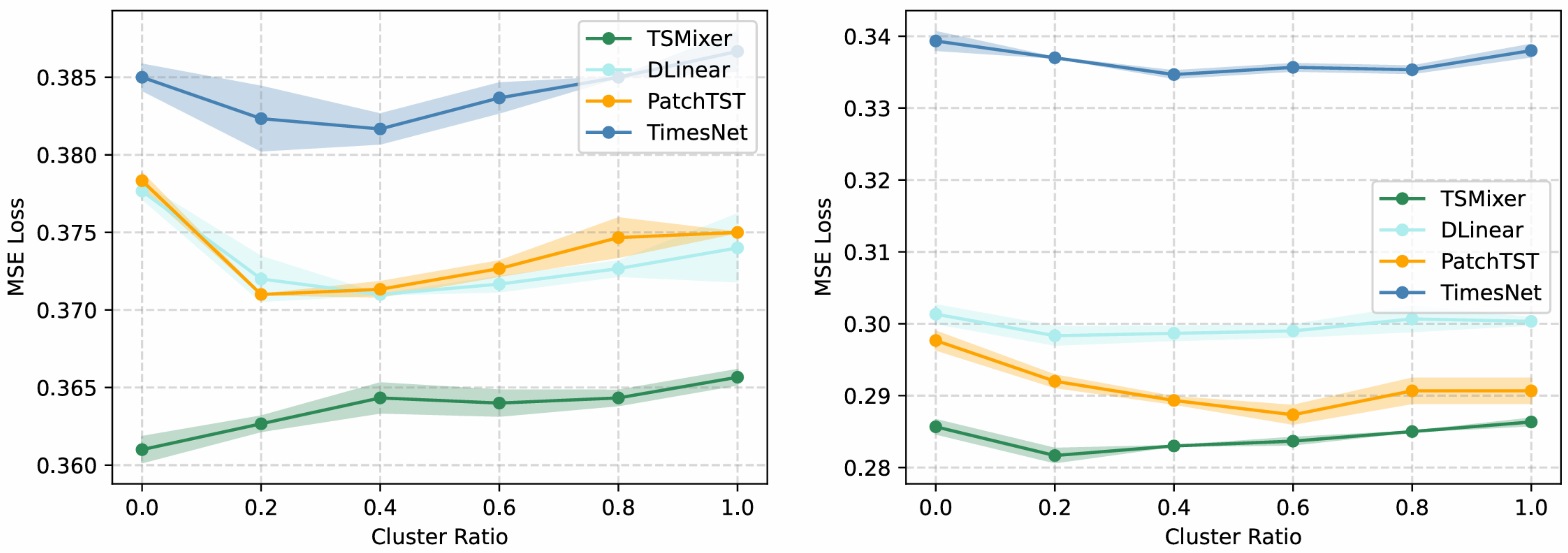


(a) ETTh1 Dataset (b) ETTm1 Dataset

05 实验：消融实验

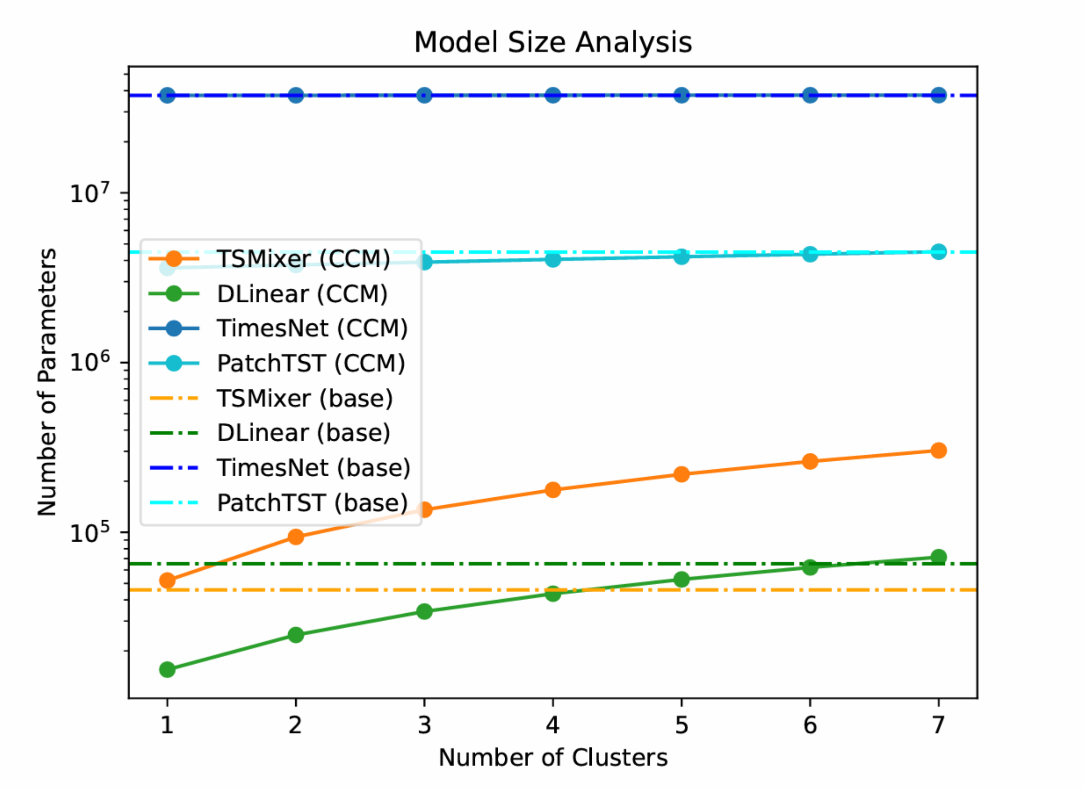
比例：聚类数量 / 通道数量

0表示所有通道放到一个聚类中，1表示所有通道各自独立

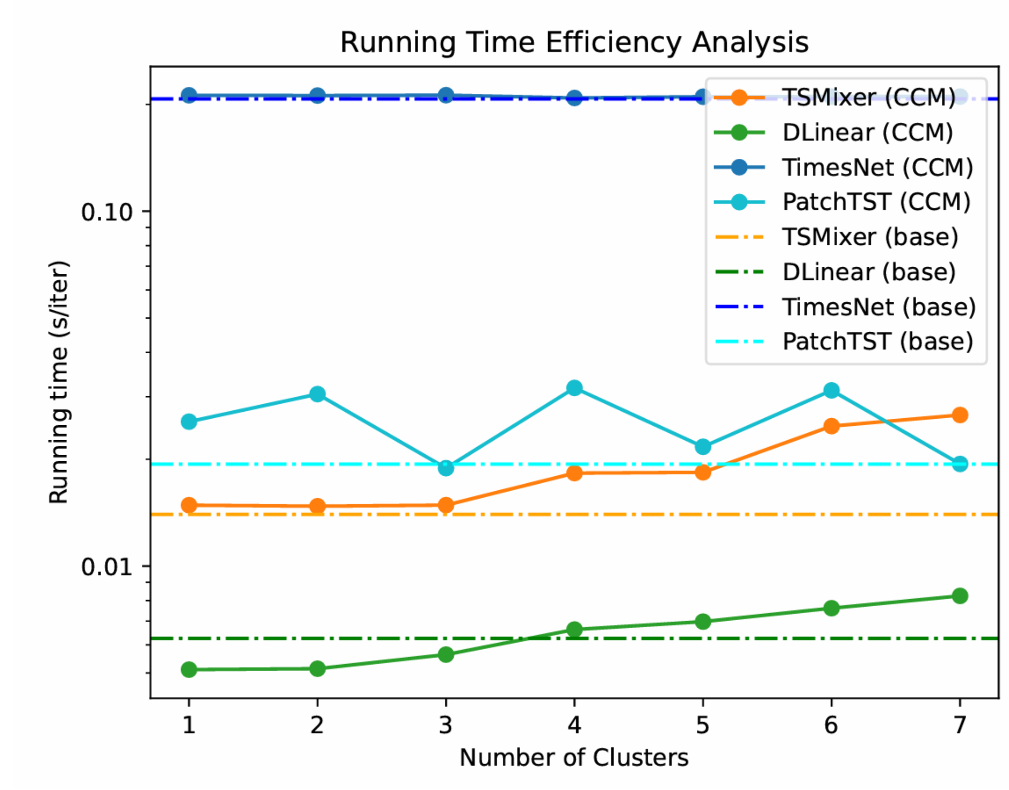


(a) MSE loss w.r.t. cluster ratios on ETTh1 dataset (b) MSE loss w.r.t. cluster ratios on ETTm1 dataset

05 实验：效率实验



(a) Model Size Analysis on ETTh1 dataset



(b) Runtime Efficiency Analysis on ETTh1 dataset

谢谢观看

MANY THANKS !

24.11.28