IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING 中科院2区
23.10.16 Early Access

ST-MAN

Spatio-**T**emporal **M**emory Augmented Multi-Level **A**ttention **N**etwork for Traffic Prediction

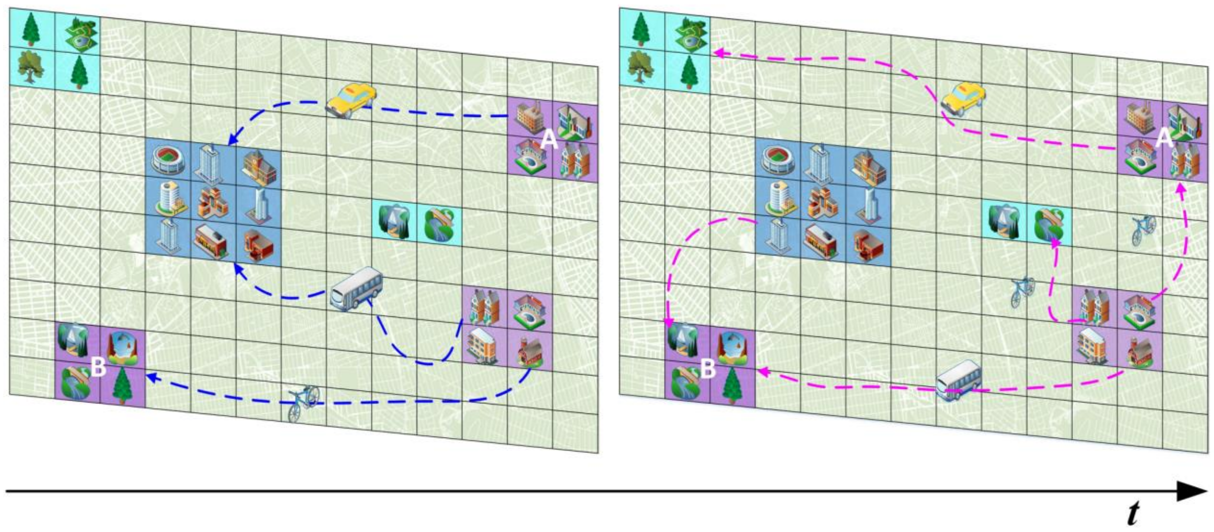
23.11.7

Presented by Yyyq

01 问题描述

➢ **long-range and long-term：远程和长期的时空因素**

- **空间依赖性**
  人群的流动性，住在郊区，在市中心上班
  具有远程语义相关性
- **时间依赖性**
  周期性模式，例如每日和每周。



02

创新点

- **时空记忆网络 STMN（Memory Networks)**
  - Key memory matrix：全局空间相关性编码作为先验知识
  - Value memory matrix：记忆编码捕获长期时间模式
- **多层注意力网络 MAN：提取 Long-range + Long-term**
  - 短期局部注意力：跨网格 + 跨区域
  - 长期全局注意力：功能区域之间的语义相关性
- **外部因素的影响**
  - 天气和日历

03

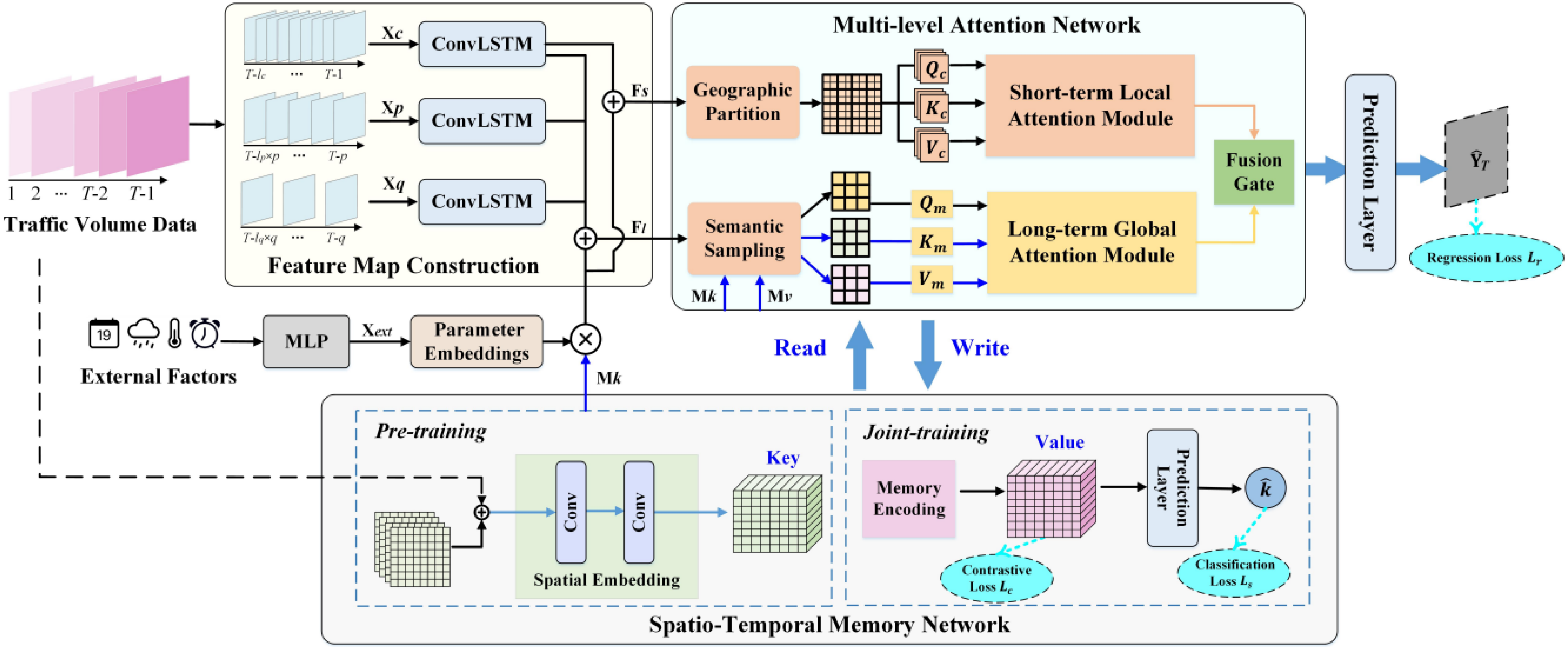
问题定义

- **网格单元（细粒度）：** H × W
- **区域（粗粒度）：** H′ × W′
- **基于网格的交通流量图**：网格在第n个时间间隔内的（移动）出行次数。

$$\mathbf{M}_t \in \mathbb{R}^{K \times H \times W}$$

- **交通流量预测**：单步预测前 T- 1 步预测第 T 步

04 算法实现：系统框架



Multi-level Attention Network
ConvLSTM
ConvLSTM
ConvLSTM
Xc
Xp
Xq
Fs
Fl
Traffic Volume Data
Feature Map Construction
Geographic Partition
Semantic Sampling
Short-term Local Attention Module
Long-term Global Attention Module
Fusion Gate
Prediction Layer
Regression Loss L_r
External Factors
MLP
Xext
Parameter Embeddings
Mk
Mv
Read
Write
Pre-training
Joint-training
Conv
Conv
Spatial Embedding
Key
Value
Memory Encoding
Prediction Layer
Contrastive Loss L_c
Classification Loss L_s
Spatio-Temporal Memory Network

04 算法实现：Feature Map Construction（数据嵌入层）

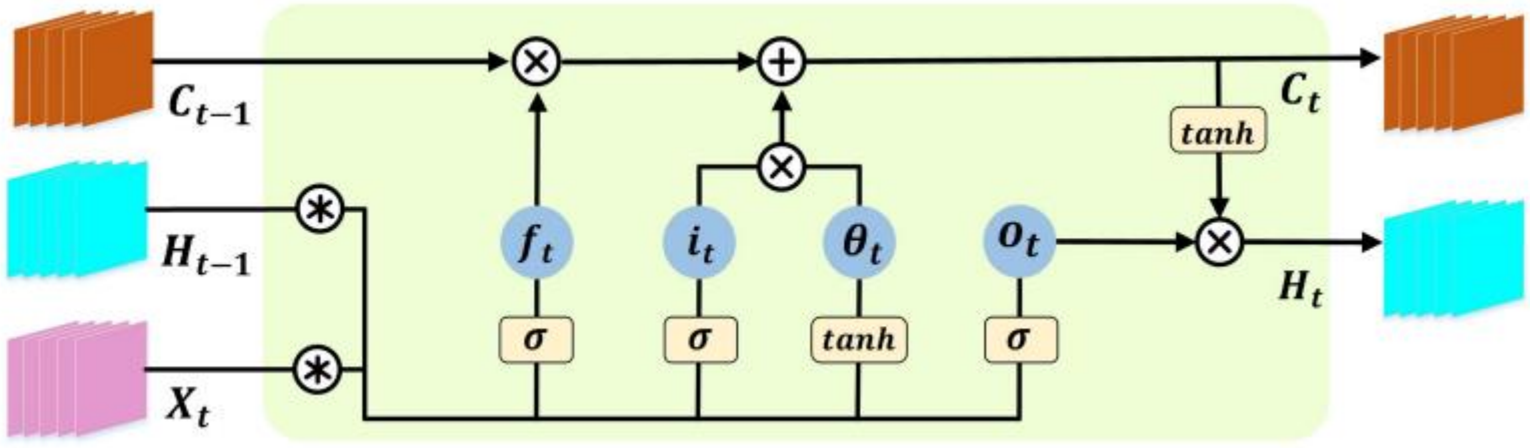
➢ **ConvLSTM**

- **前 T – 1 个时间步：**最近、日、周

$$\mathbf{X}_c = [\mathbf{M}_{T-l_c}, \mathbf{M}_{T-(l_c-1)}, \ldots, \mathbf{M}_{T-1}] \in \mathbb{R}^{Kl_c \times H \times W}$$

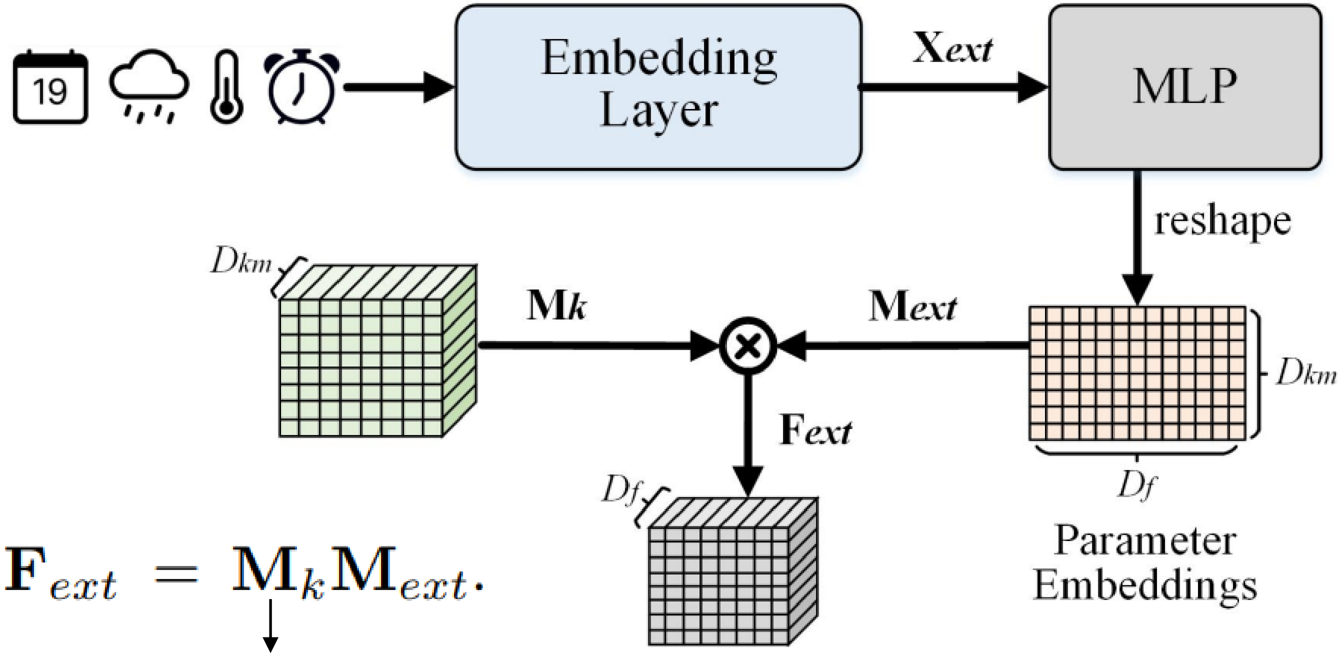
$$\mathbf{X}_p = [\mathbf{M}_{T-l_p \times p}, \mathbf{M}_{T-(l_p-1)\times p}, \ldots, \mathbf{M}_{T-p}] \in \mathbb{R}^{Kl_p \times H \times W},$$

$$\mathbf{X}_q = [\mathbf{M}_{T-l_q \times q}, \mathbf{M}_{T-(l_q-1)\times q}, \ldots, \mathbf{M}_{T-q}] \in \mathbb{R}^{Kl_q \times H \times W},$$



04 算法实现：Feature Map Construction（数据嵌入层）

➢ **外部因素：** 对不同网格单元的位置感知影响

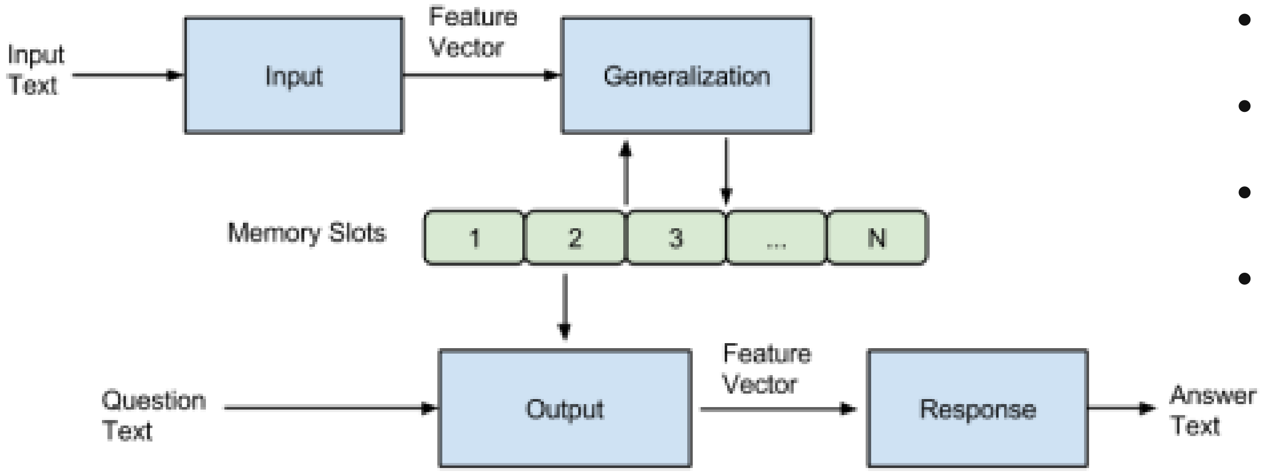


$$\mathbf{F}_{ext} = \mathbf{M}_k \mathbf{M}_{ext}.$$

Key matrix

算法实现：时空记忆网络 STMN（数据编码层）

- **Memory networks**
  - 一种可读写的外部记忆模块
  - 用长期记忆（long-term memory）来保存问答的知识或者聊天的语境信息



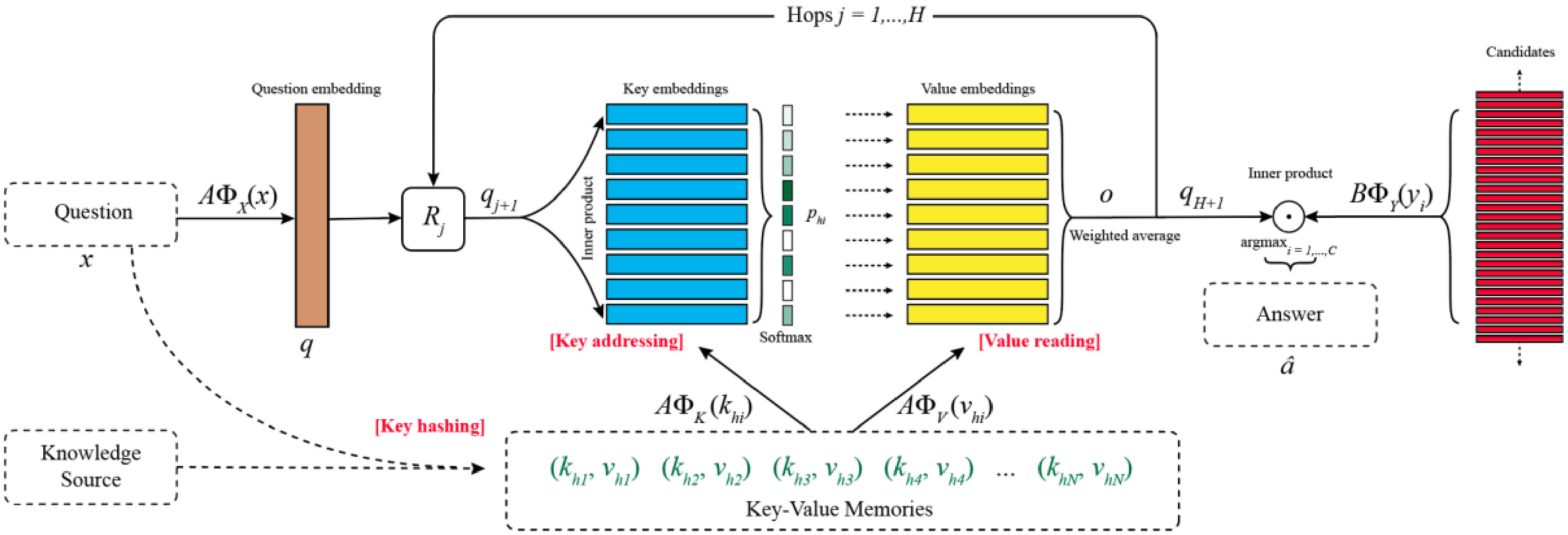
- I ：输入的文本→特征向量
- G ：用新的输入数据更新 memories
- O ：根据Question对memory的内容进行权重处理
- R ：将输出转化为自然语言

[1] Weston J, Chopra S, Bordes A. Memory networks[C]//3rd International Conference on Learning Representations, ICLR 2015. （ ICLR 2015，Facebook AI Research）

04 算法实现：时空记忆网络 STMN（数据编码层）

➢ Key-Value Memory Networks

- 对先验知识（阅读的信息）进行编码
- 以key-value对的形式编码进记忆网络（key负责寻址，value负责读取）

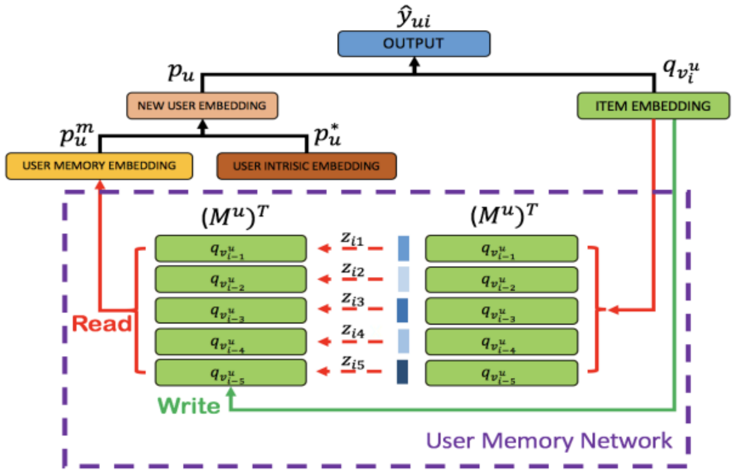


[2] Miller A, Fisch A, Dodge J, et al. Key-Value Memory Networks for Directly Reading Documents[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016: 1400-1409. （ACL2016，Facebook AI Research）

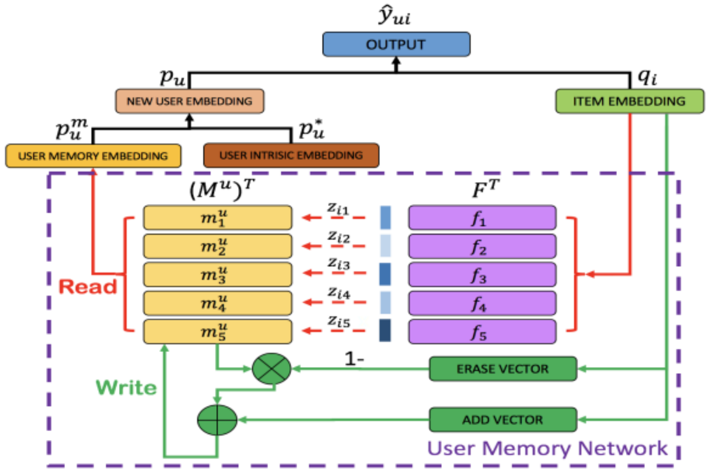
04 算法实现：时空记忆网络 STMN（数据编码层）

➢ **Sequential Recommend with User Memory Networks**

- 为每个用户使用外部记忆单元，记忆历史行为
- 预测时，动态注意力读出记忆 作为用户Embedding表示
- 注意力机制衡量历史行为对当前物品 推荐的重要性



(a) Item-level RUM



(b) Feature-level RUM

[3] Xu Chen, Hongteng Xu, Yongfeng Zhang, Jiaxi Tang, Yixin Cao, Zheng Qin, and Hongyuan Zha. 2018. Sequential Recommendation with User Memory Networks. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM '18). (School of Software Tsinghua University)

04

算法实现：时空记忆网络 STMN（数据编码层）

- **Spatio-Temporal Memory Network**
  - 在本文中的目的：引入先验知识（时空信息）

① 空间信息嵌入

Word2vec编码网格为向量

空间分布：曼哈顿距离 $d(i,j)=|x_i-x_j|+|y_i-y_j|$

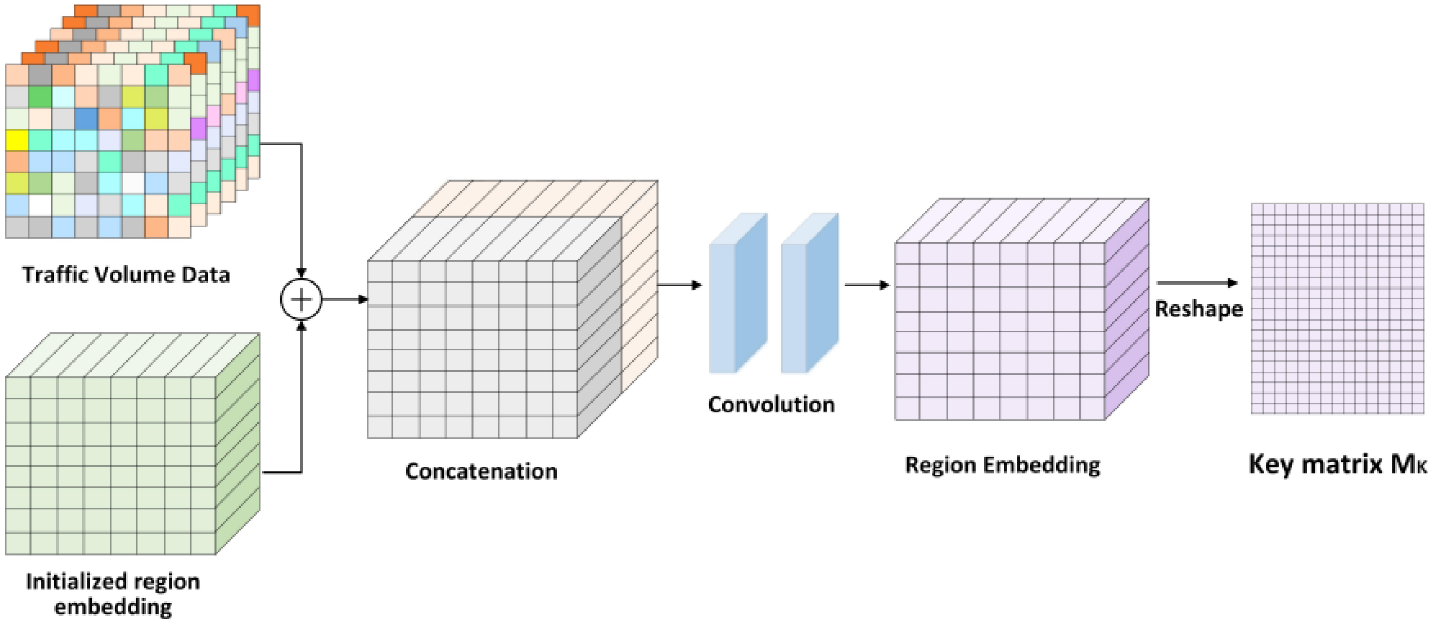
功能相似性：Pearson相关系数 $\rho_{X,Y}=\frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y}=\frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y},$

04 算法实现：时空记忆网络 STMN（数据编码层）

① 空间信息嵌入

Word2vec编码网格为向量

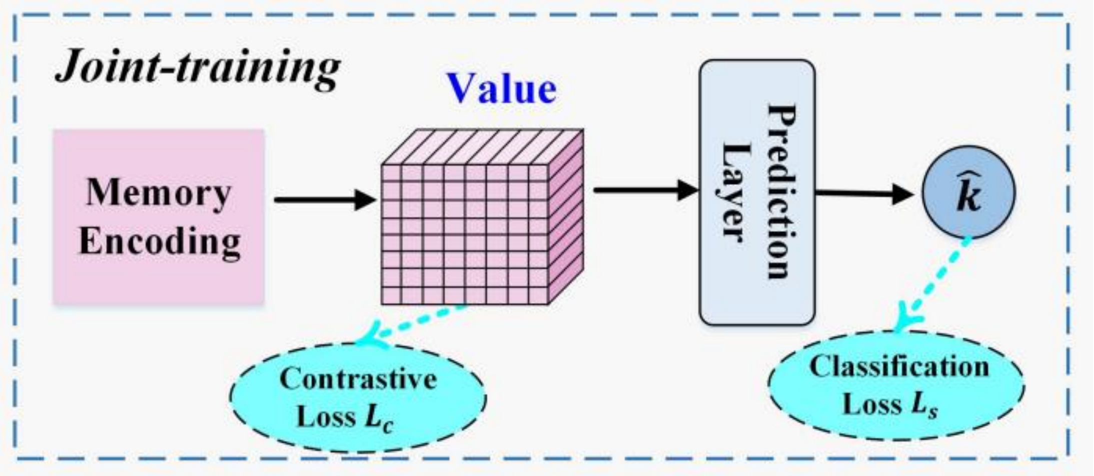
- **空间分布：** 曼哈顿距离 $$d(i,j) = |x_i - x_j| + |y_i - y_j|$$
- **功能相似性：** Pearson相关系数 $$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y},$$



04 算法实现：时空记忆网络 STMN（数据编码层）

② 记忆编码学习时间模式

基于Supervised Contrastive Learning



- **聚类**：K-means++将所有网格聚成S个类别

- **区分正负样例**：

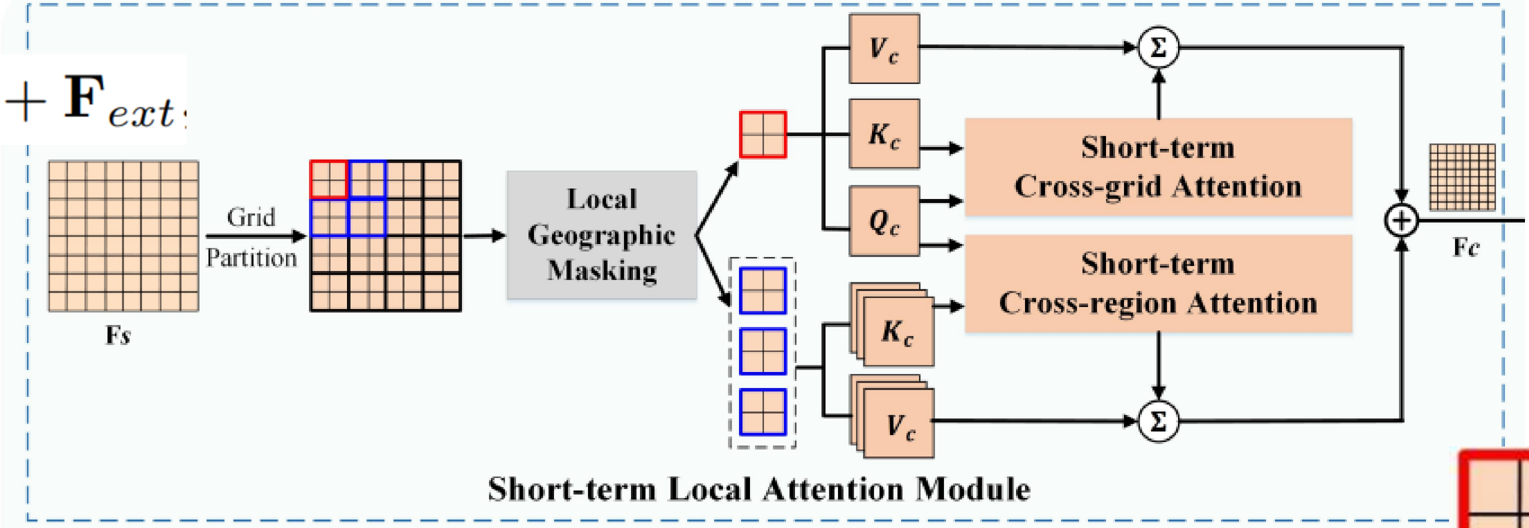
$$B^1 = (g_1^1, \ldots g_i^1 \ldots, g_S^1),$$

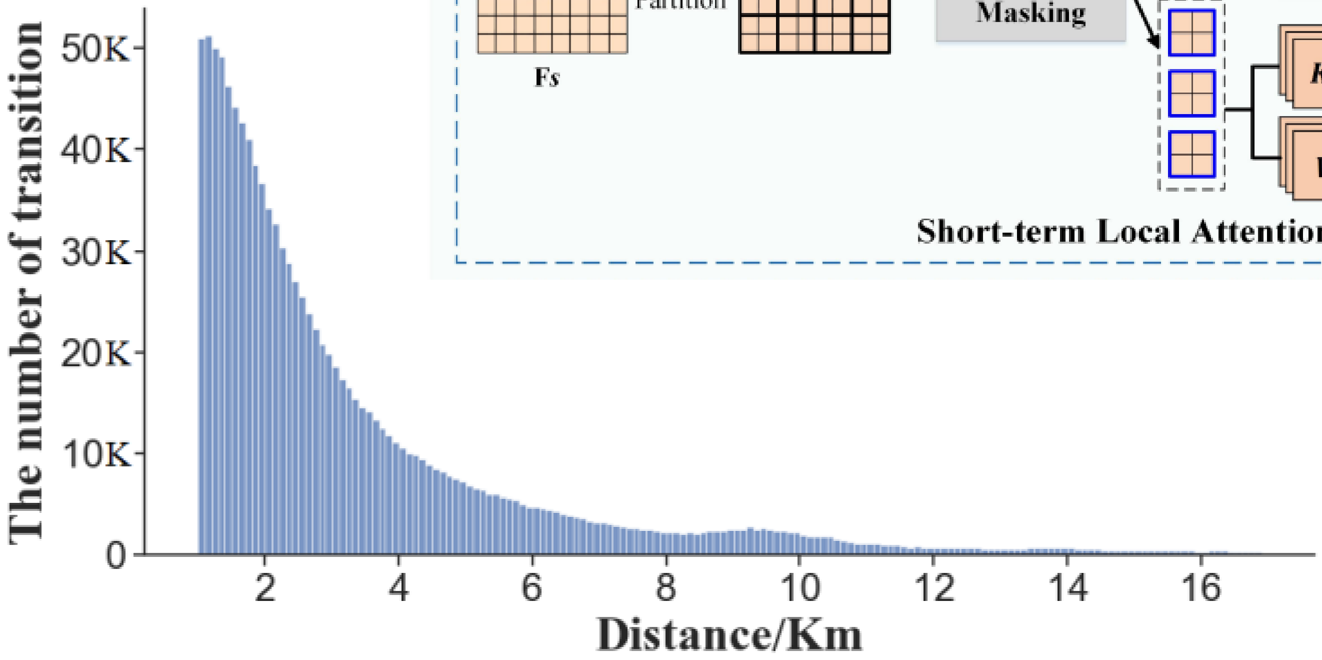
$$B^2 = (g_1^2, \ldots g_i^2 \ldots, g_S^2)$$

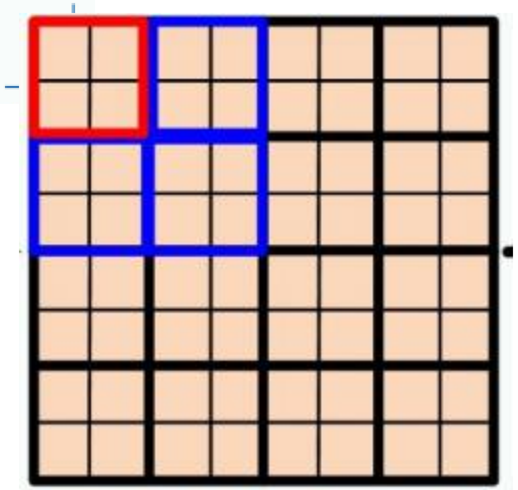
04 算法实现：Multi-Level Attention Network

➢ **短期局部注意SLA：**相邻网格/区域，近邻时间步 + 外部因素

$$\mathbf{F}_s = \mathbf{F}_c + \mathbf{F}_{ext}$$







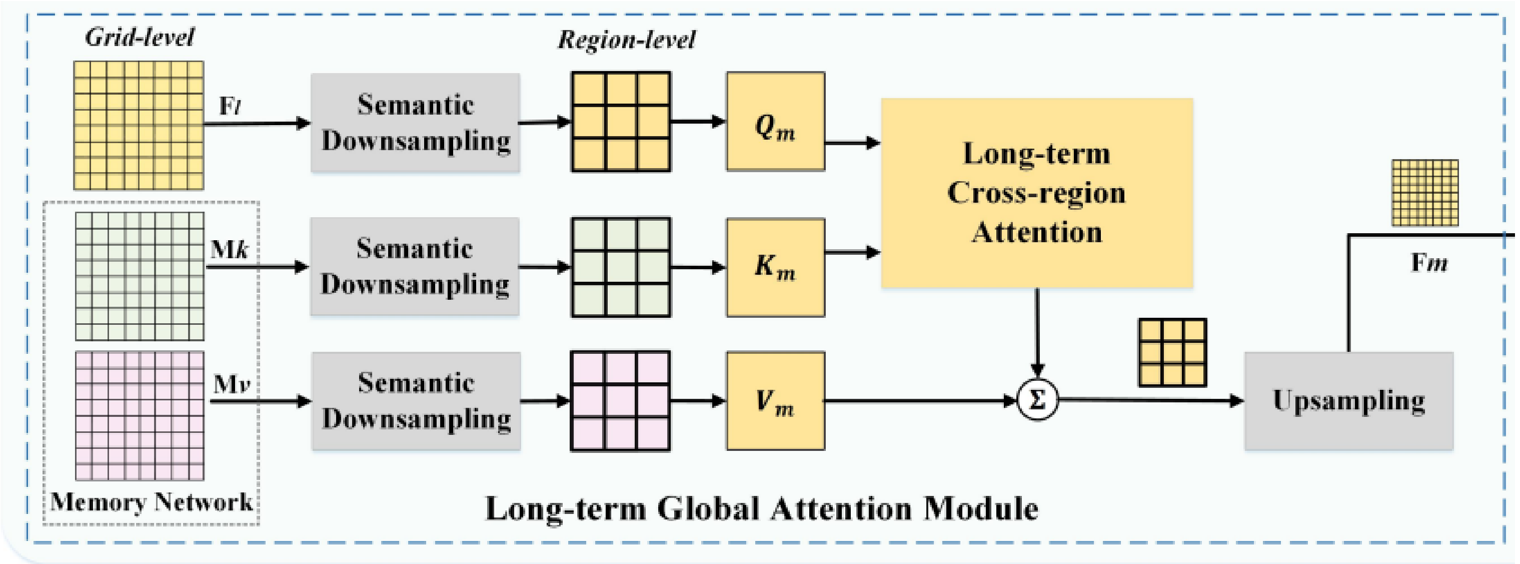
04 算法实现：Multi-Level Attention Network

➤ **长期全局注意LGA：**全局语义信息，近邻/日/周时间步 + 外部因素

$$\mathbf{F}_l = \mathbf{F}_c + \mathbf{F}_p + \mathbf{F}_q + \mathbf{F}_{ext}$$

key memory matrix $\mathbf{M}_k$

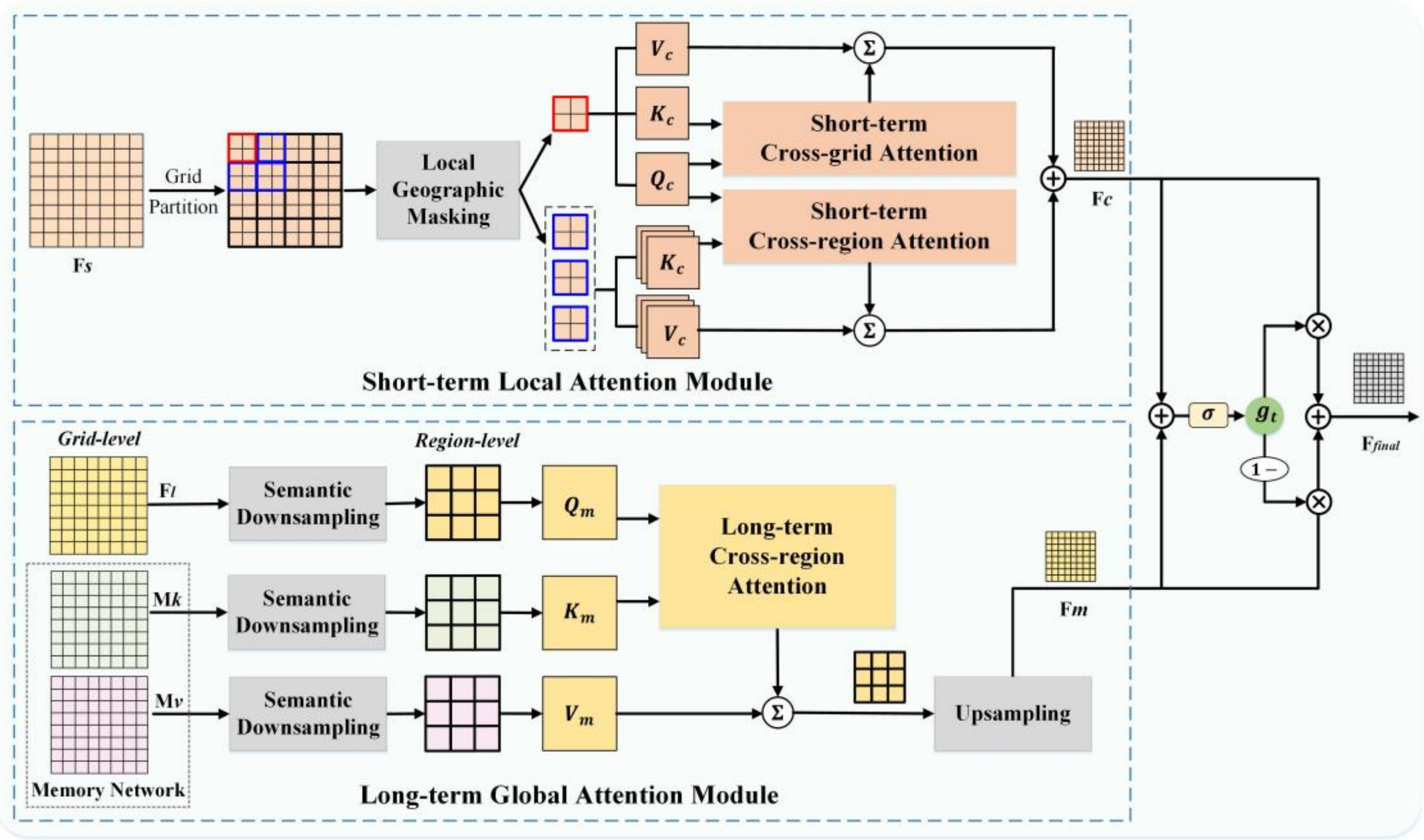
value memory matrix $\mathbf{M}_v$



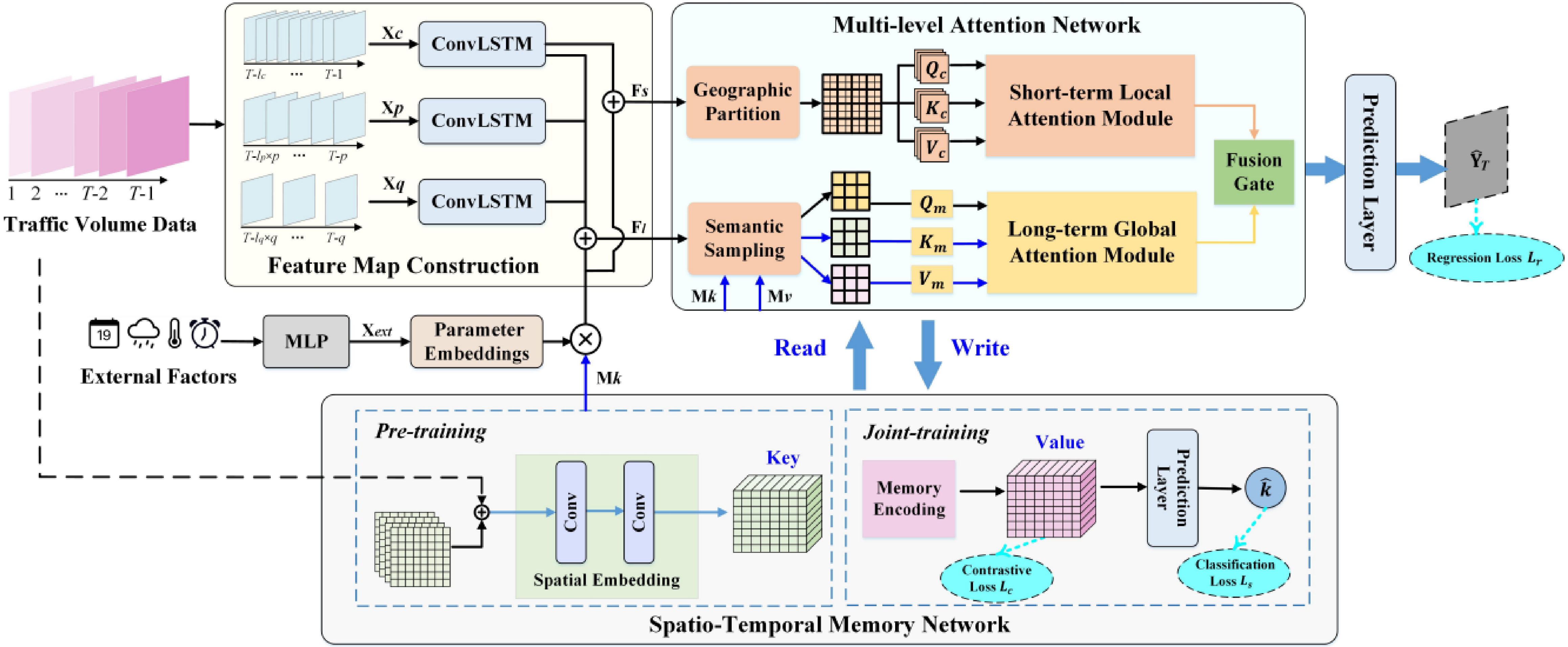
网格级 128×128 —卷积→ 区域级 32×32 → —N²-N规范化→ 网格级 128×128

04 算法实现：Multi-Level Attention Network

➢ **Fusion机制：**门控机制



04 算法实现：系统框架



Multi-level Attention Network
ConvLSTM
ConvLSTM
ConvLSTM
Xc
Xp
Xq
Fs
Fl
Geographic Partition
Semantic Sampling
Short-term Local Attention Module
Long-term Global Attention Module
Fusion Gate
Prediction Layer
Regression Loss L_r
Traffic Volume Data
Feature Map Construction
External Factors
MLP
Xext
Parameter Embeddings
Mk
Mv
Read
Write
Pre-training
Spatial Embedding
Conv
Key
Joint-training
Memory Encoding
Value
Contrastive Loss L_c
Classification Loss L_s
Spatio-Temporal Memory Network

05

实验：数据集

Dataset	TaxiNYC	TaxiDC	BikeNYC	BikeDC	TaxiBJ+
Grids	10×20	16×16	10×20	16×16	128×128
Start time	1/1/2015	5/1/2015	1/1/2017	1/1/2017	7/1/2013
End time	7/1/2015	1/1/2016	12/31/2017	12/31/2017	10/30/2013
Time interval	1 hour	1 hour	1 hour	1 hour	30 minutes
External factors	/	/	/	/	7

区域：5×10

区域：16×16

区域：8×8

05

实验：对比实验

Datasets	TaxiNYC		TaxiDC		BikeNYC		BikeDC		TaxiBJ+	
Metrics	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
HA	10.0573	5.1072	33.6695	14.1543	30.8275	15.6316	12.445	7.0751	26.5702	18.8546
ARIMA	9.0998	5.0213	28.6089	10.7037	23.4854	11.6232	10.7253	6.4053	22.8099	15.7216
LR	7.4235	4.5586	13.0652	6.4329	15.2656	7.9387	8.1601	4.5318	15.4554	9.9221
Tree	8.8234	5.0987	14.7149	7.0631	16.1438	8.4188	8.0506	4.8498	16.7589	10.8695
MLP	9.8668	5.4152	19.2622	7.9245	11.8643	6.2067	5.5737	3.4883	9.4351	4.6531
CNN	5.1658	2.4539	5.628	1.8363	7.2354	3.4372	2.2232	0.8729	6.7147	3.0062
FC-LSTM	5.6171	2.6308	6.5573	2.0517	8.5537	3.249	2.6236	0.8415	6.961	3.2109
ConvLSTM	4.5731	2.0385	4.9085	1.4229	5.8726	2.8113	1.6271	0.5915	6.2903	3.0024
DMVST-Net	5.3104	2.7974	6.0765	2.666	6.4303	2.7589	1.7522	0.6548	6.6276	3.1431
DeepST	4.5453	2.0262	5.624	1.8486	5.5124	2.715	1.6292	0.6833	6.6942	3.1975
ST-ResNet	4.4784	2.3474	5.5259	1.8994	5.3987	2.6863	1.4385	0.5343	6.6561	3.1331
STDN	4.3754	1.9862	5.3398	1.6181	5.4237	2.7146	1.4962	0.6032	6.4877	3.0007
DeepSTN+	4.2504	2.2154	5.0929	1.8931	5.0076	2.3536	1.2872	0.4872	5.6976	3.0612
SAConvLSTM	4.2646	1.8785	4.9074	1.441	5.1384	2.4891	1.3471	0.5108	5.5434	2.7641
PDFormer	4.2369	1.9475	4.8228	1.5389	4.9562	2.3047	1.2939	0.4903	5.5075	2.6828
ST-MAN	**4.1396**	**1.8272**	**4.5205**	**1.4052**	**4.8264**	**2.1524**	**1.2397**	**0.4751**	**5.3876**	**2.5264**

实验：消融实验

Ablation Analysis on Five Datasets

Datasets	TaxiNYC		TaxiDC		BikeNYC		BikeDC		TaxiBJ+	
Metrics	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ST-MAN w/o ConvLSTM	4.1952	2.0028	5.5153	1.7381	5.0138	2.3182	1.5446	0.6792	5.6374	2.8183
ST-MAN w/o STMN	4.4257	2.2412	5.3895	1.7637	5.2064	2.3236	1.5732	0.5643	5.6694	2.9095
ST-MAN w/o MAN	4.1881	1.9269	5.0817	1.5258	4.9607	2.2017	1.3856	**0.4732**	5.8325	3.0816
ST-MAN	**4.1396**	**1.8272**	**4.5205**	**1.4052**	**4.8264**	**2.1524**	**1.2397**	0.4751	**5.3876**	**2.5264**

Effect of Attention Modules on Five Datasets

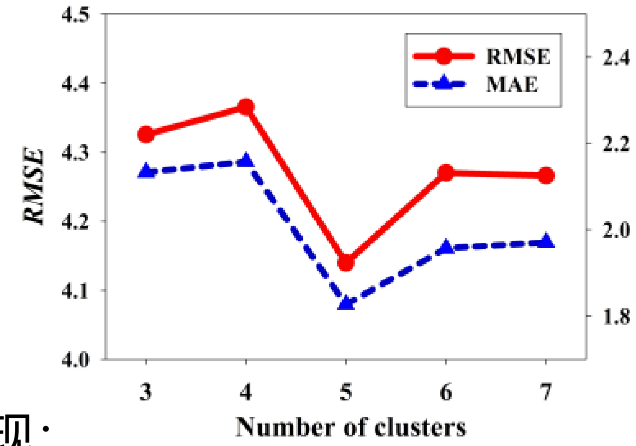
Datasets	TaxiNYC		TaxiDC		BikeNYC		BikeDC		TaxiBJ+	
Metrics	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ST-MAN w/o SGLA	4.1556	2.0133	4.9256	1.6048	4.9819	2.223	1.2975	0.4923	5.7185	3.0278
ST-MAN w/o SRLA	4.1803	2.0849	5.1267	1.6163	5.0736	2.275	1.3256	0.5162	5.9274	3.1246
ST-MAN w/o LGA	4.4257	2.2412	5.3895	1.7637	5.2064	2.3236	1.5732	0.5643	5.6694	2.9095
ST-MAN	**4.1396**	**1.8272**	**4.5205**	**1.4052**	**4.8264**	**2.1524**	**1.2397**	**0.4751**	**5.3876**	**2.5264**

实验：聚类个数

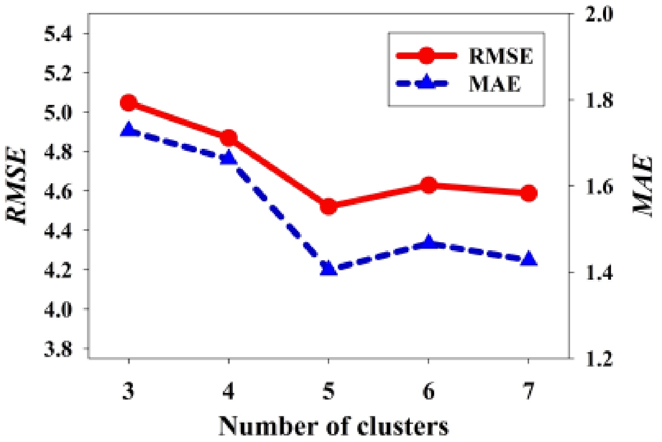
聚类个数 = 5

城市的功能区趋于稳定，不会短期消失或出现；

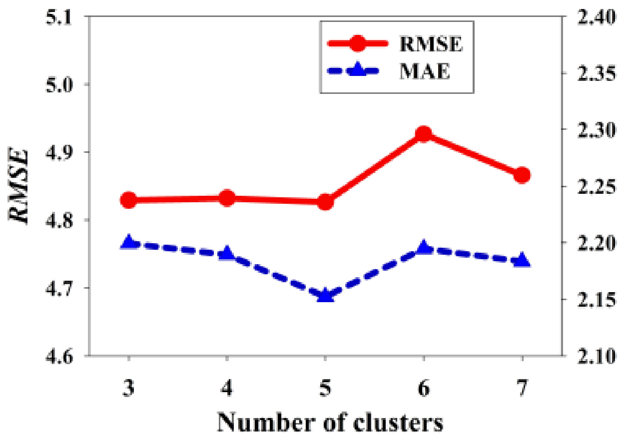
（商业区、居住区、行政区、学院区、风景区）



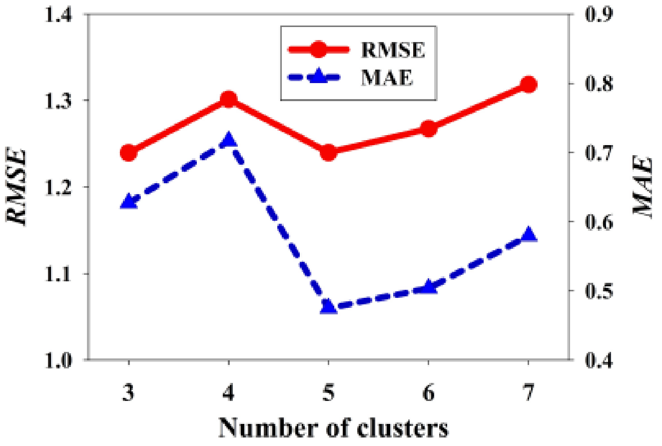
(a) TaxiNYC



(b) TaxiDC



(c) BikeNYC



(d) BikeDC

06 总结：个人体会

- **别样的拼凑，像是在写学习总结**
  - 主要模型架构其实还是transformer，但通篇没有明提transformer
  - 嵌入层：简单的卷积 → 成熟的模型 ConvLSTM
  - 基本把现在常见的改进思路全都杂糅在一个模型里面
    - 空间：地理邻域 + 全局语义
    - 时间模式：聚类
  - 每一个改进思路再堆砌一个高大上技术名词
- **缺点**
  - 语言组织太差了，基本属于车轱辘话来回说
  - **问题描述**，基本没有谈实际应用的痛点，直接就是技术难点
  - **实验**做的偏简单了，前面说引入记忆网络是为了高参数量带来的运算压力和特征转换造成的信息丢失，但并没有辅助实验去验证效率这方面
  - 公式和模型图上符号还存在小错误

谢谢观看

MANY THANKS !

23.11.7