Multi-modal Time Series Analysis: A Tutorial and Survey

2025/04/13

I 引言

什么是多模态时间序列分析？

- 结合时间序列数据与其他补充模态（如文本、图像等）进行分析。

为什么重要？

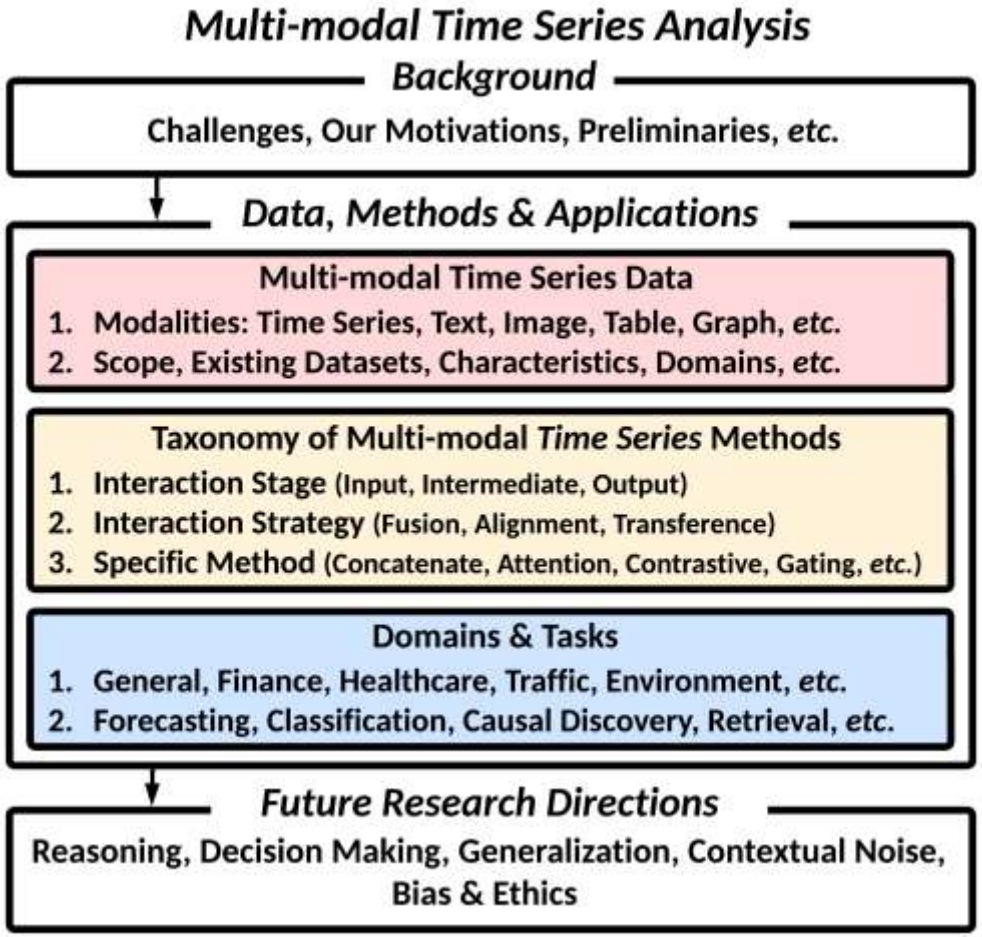
- 数据来源日益多样化(文本、图像、表格等)
- 多模态提供更丰富的信息，显著提升下游任务（如预测、分类）的性能

核心挑战

- 数据异构性（不同统计特性、结构、维度）
- 时间不对齐（不同时间戳或粒度）。
- 固有的噪声和不相关信息。

综述目标：

- 提供多模态时间序列数据集和方法的系统性、最新概述
- 提出一个统一的跨模态交互框架。



II 多模态时间序列数据

常见模态

- 时间序列：核心数据
- 表格（Tabular）：事件日志、人口统计信息等。
- 文本（Text）：临床记录、新闻、社交媒体帖子等。
- 图像（Image）：医学影像、卫星图像等。
- 图（Graph）：社交网络关系、交通网络结构等。
- 音频（Audio）：可视为特殊类型的时间序列。

代表性数据集：

Domain	**Dataset** (Superscripts include the URLs to the datasets)	**Modalities**
Healthcare	MIMIC-III [35][1], MIMIC-IV [34][2] ICBHI [65][3], Coswara [4][4], KAUH [21][5], PTB-XL [71][6], ZuCo [14, 26][7] Image-EEG [22][8]	TS, Text, Tabular TS, Text TS, Image
Finance	FNSPID [17][9], ACL18 [84][10], CIKM18 [79][11], DOW30 [11][12]	TS, Text
Multi-domain	Time-MMD [53][13], TimeCAP [42][14], NewsForecast [73][15], TTC [37][16], CiK [77][17], TSQA [38][18]	TS, Text
Retail	VISUELLE [70][19]	TS, Image, Text
IoT	LEMMA-RCA [40][20]	TS, Text
Speech	LRS3 [1][21], VoxCeleb2 [13][22]	TS (Audio), Image
Traffic	NYC-taxi, NYC-bike [48][23]	ST, Text
Environment	Terra [10][24]	ST, Text

III Cross-modal Interactions with Time Series

III 跨模态交互框架

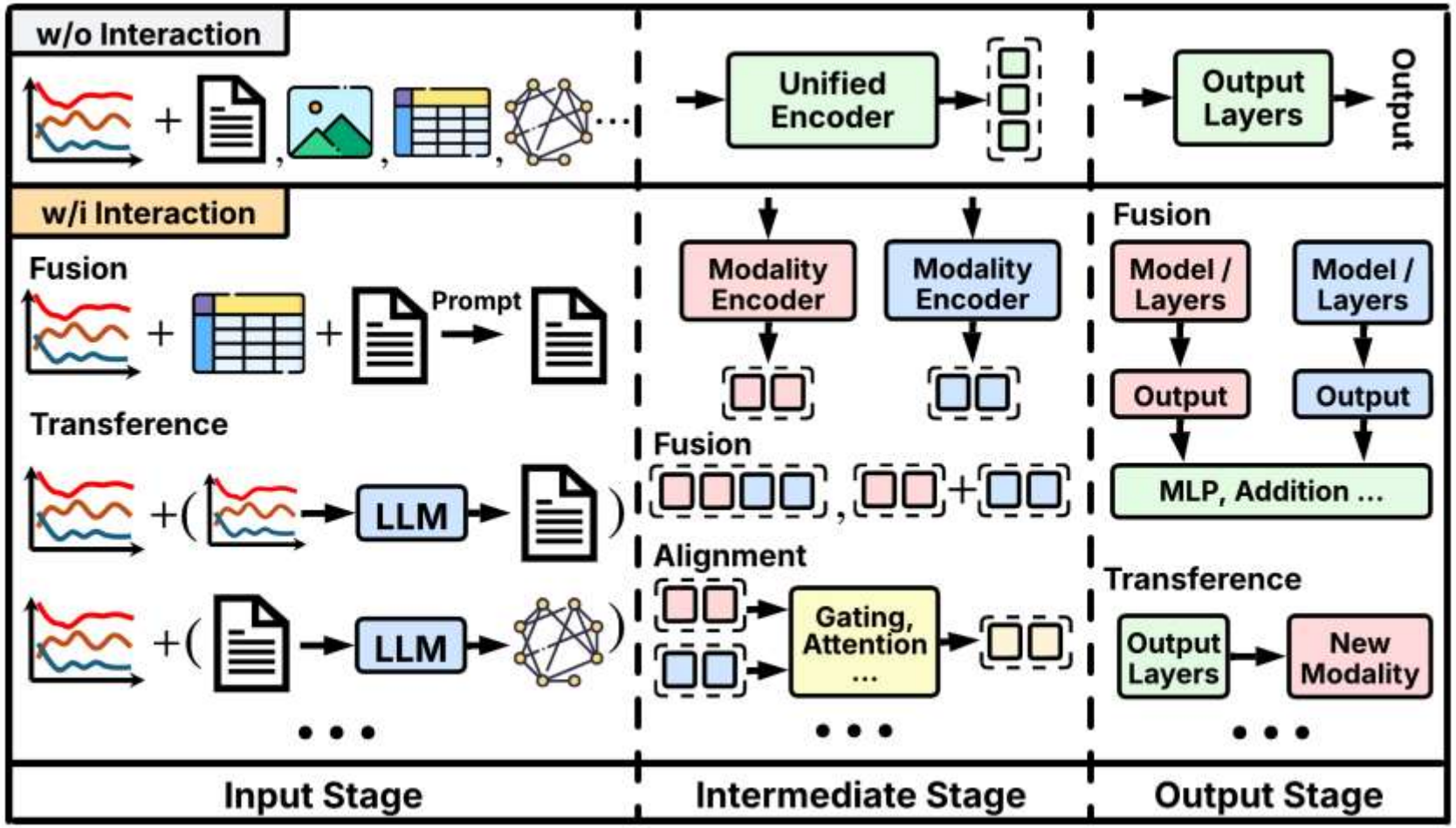
核心思想：根据模态间交互的方式和阶段对现有方法进行分类

三种交互方式

- 融合（Fusion）：整合异构模态以捕捉互补信息
- 对齐（Alignment）：确保模态间关系的保留和语义一致性
- 转换（Transference）：在不同模态之间进行映射（如生成、翻译）

三个交互阶段：

- 输入阶段
- 中间阶段（表示层或中间输出）
- 输出阶段



III 交互方式 – 融合

目标： 整合不同模态信息以改进时间序列建模。

输入阶段融合

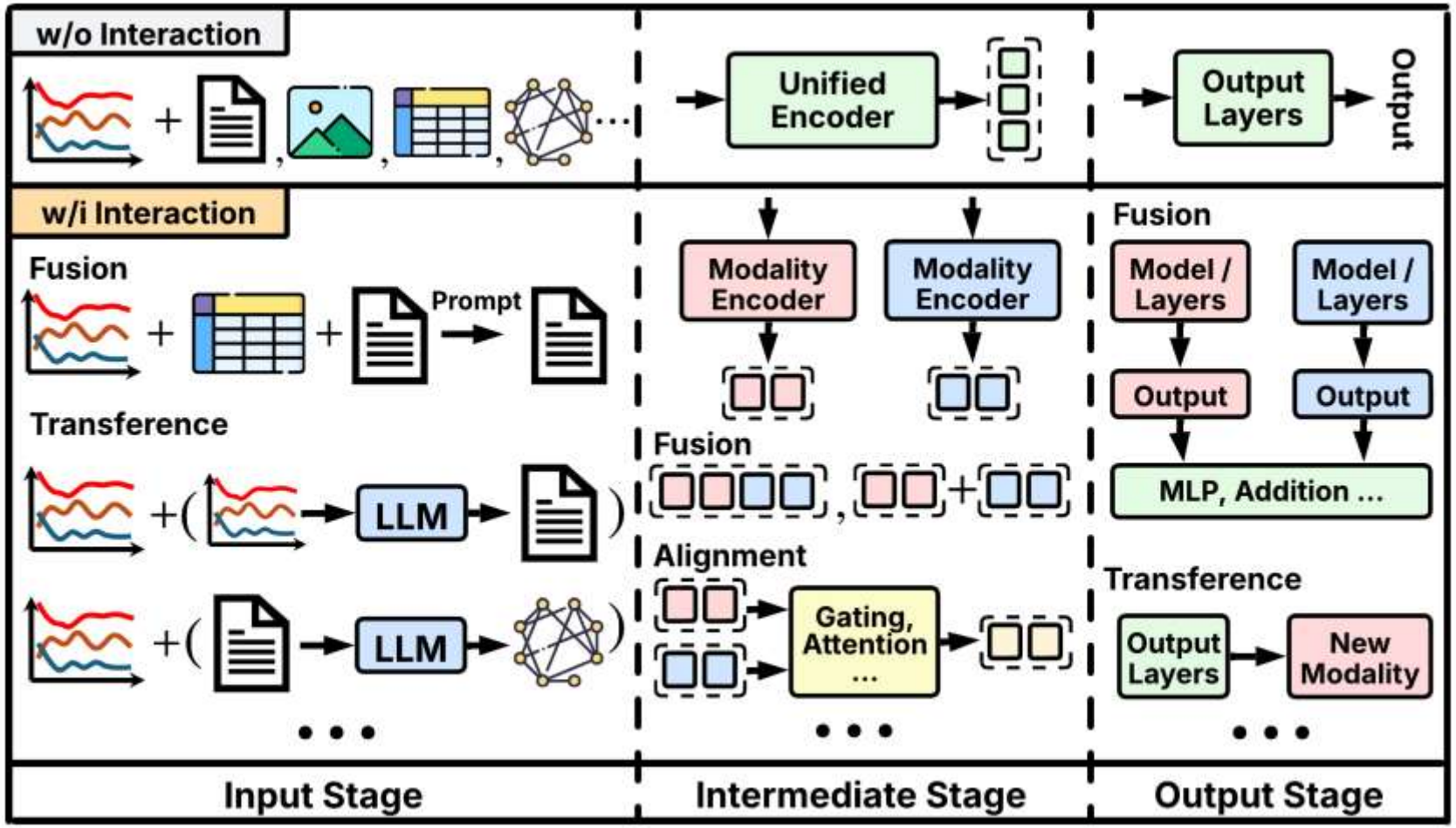
- 将TS、表格、文本整合为统一的文本提示(prompt)，输入LLM
- 将配对的文本嵌入作为TS的附加变量

中间阶段融合

- 常用方法：加法、拼接多模态表示

输出阶段融合

- 来自不同模态的预测或模型的输出在输出阶段融合



III 交互方式 – 对齐

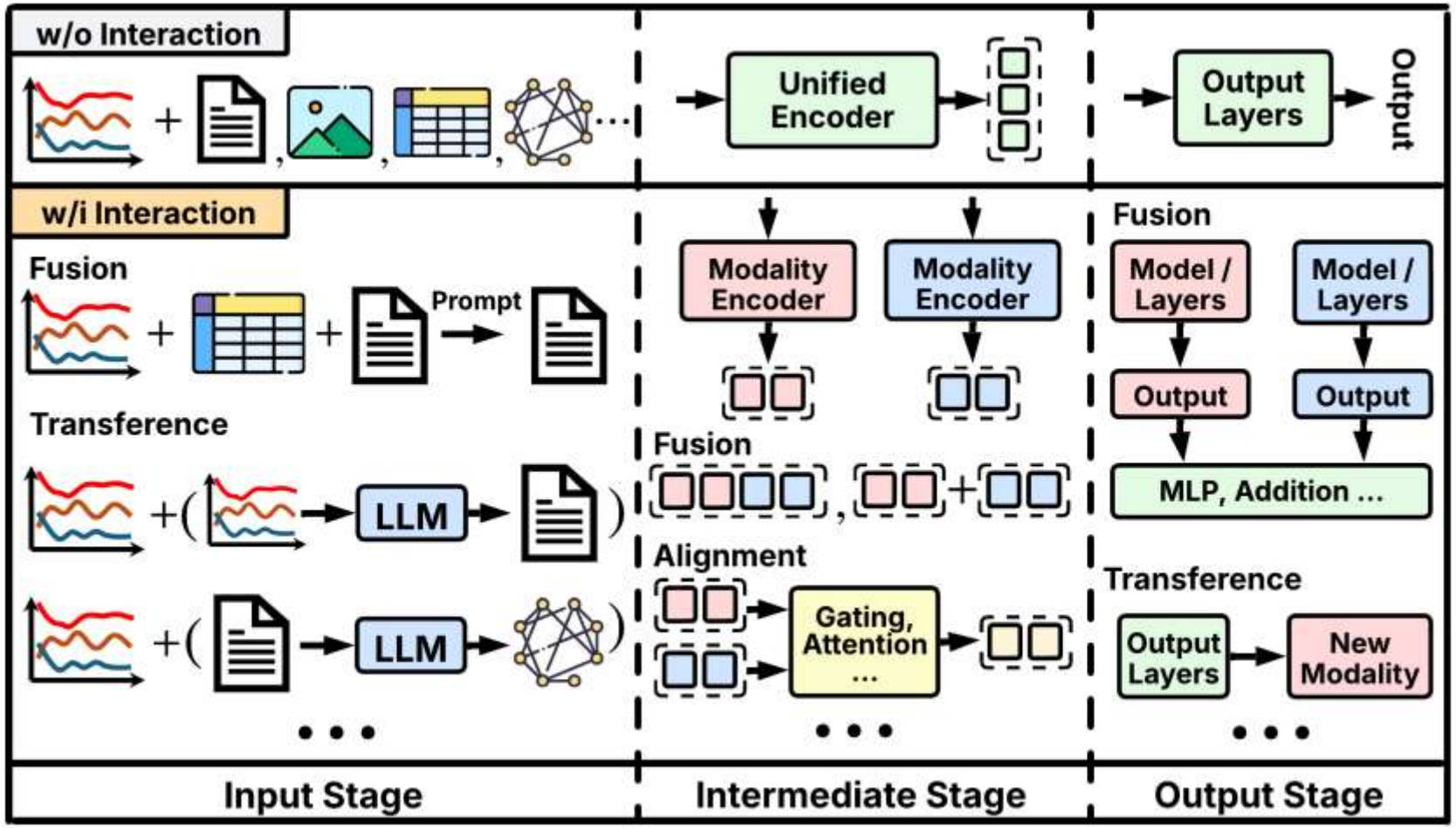
目标：确保不同模态在统一学习框架中保留关系和语义连贯。解决模态间的异质性、时间不一致性以及噪声问题

输入阶段对齐

- 数据预处理：解决时间不对齐、采样间隔/粒度差异问题。

中间阶段对齐

- 自注意力：联合、无向对齐所有模态。
- 交叉注意力：以时序为查询Q，其他模态作为K和V，有向对齐
- 门控机制：参数化过滤，调节各模态影响
- 图模态的对齐：利用图结构进行对齐（如图卷积）



III 交互方式 – 转换

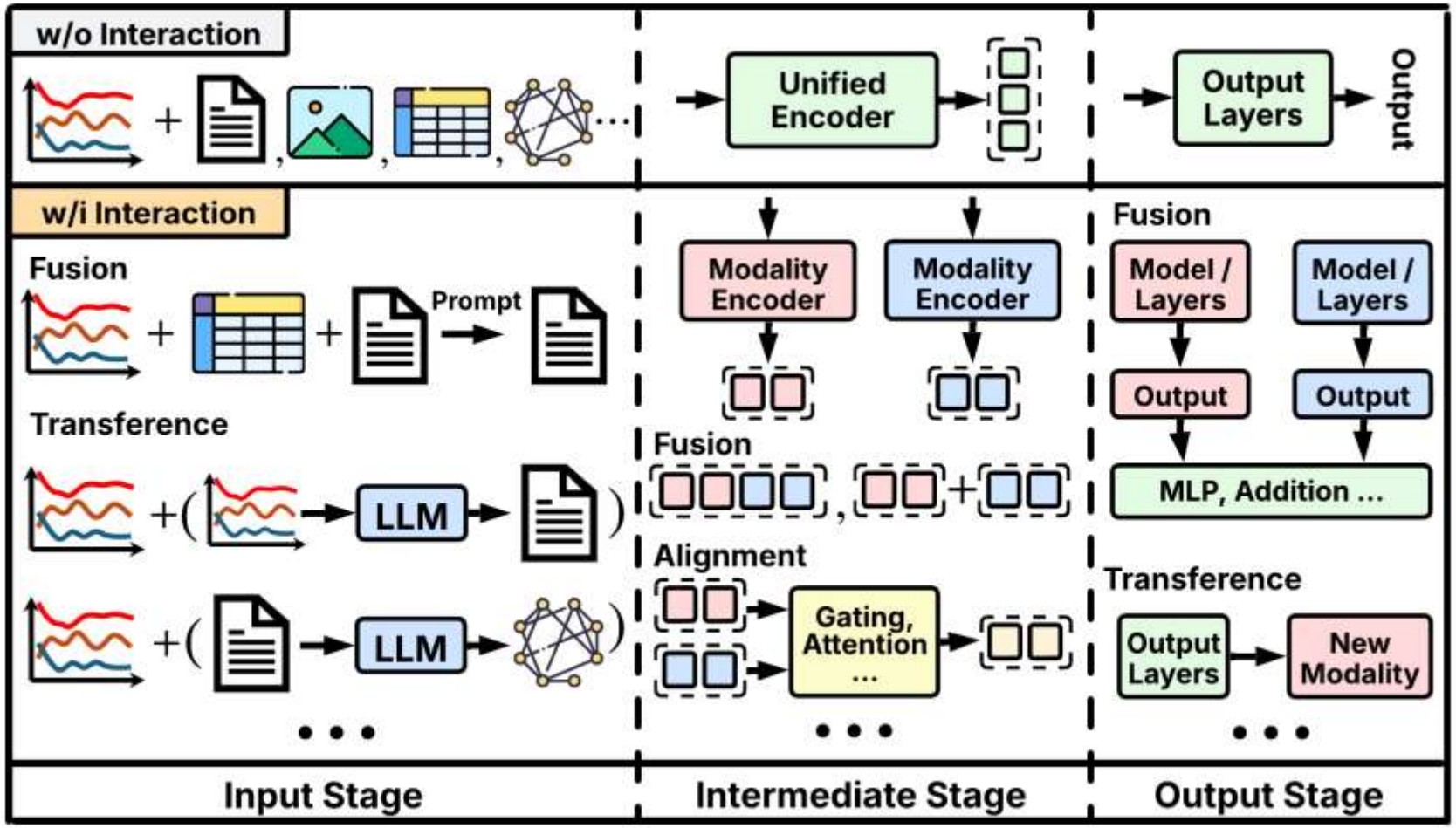
定义： 在不同模态间进行映射（推断、合成）

输入阶段转换(主要用于模态增强)

- 丰富样本，提供替代表示
- 示例：使用元信息生成文本描述 ；利用LLM生成文本/图；将TS转换为图像或表格
- 用途：作为语义锚点指导表示对齐；为LLM提供上下文

中间/输出阶段融合(任务导向)

- 中间阶段：生成待优化的初始解
- 输出阶段：新模态生成（如基于文本的时间序列数据检索）



III 方法分类

系统地分类超过40种代表性的多模态时间序列方法

- **方法 (Method)**
- **模态 (Modality)**：明确指出每种方法处理的数据类型
- **应用领域 (Domain)**：如通用、金融、交通等
- 任务 (Task)：如预测、分类、检索、生成等。
- **跨模态交互(Cross-Modal Interaction)**：交互发生的阶段 (Stage)（输入/中间/输出），以及采用的交互策略 (Fusion, Alignment, Transference)。具体的实现技术 (Method)，比如：拼接(Concat)、加法(Addition)、自注意力(Self-attention)、交叉注意力(Cross-attention)
- **大模型 (Large Model)**：标注是否采用大型预训练模型

	Method	Modality	Domain	Task	Cross-Modal Interaction					Large Model	Year	Code
					Stage	Fusion	Align.	Trans.	Method			
通用领域	Time-MMD [33]	TS, Text	General	Forecasting	Output	✓	✗	✗	Addition	Multiple	2024	Yes[1]
	Wang et al. [73]	TS, Text	General	Forecasting	Input Intermediate	✓ ✓	✗ ✓	✗ ✗	Prompt Prompt; LLM Reasoning	LLaMa2 GPT-4 Turbo	2024	Yes[2]
	GPT4MTS [30]	TS, Text	General	Forecasting	Intermediate	✓	✓	✗	Addition; Self-attention	GPT-2	2024	No
	TimeCMA [31]	TS, Text	General	Forecasting	Input Intermediate	✗ ✓	✗ ✓	✓ ✗	Meta-description Addition; Cross-attention	GPT-2	2025	Yes[3]
	MOAT [41]	TS, Text	General	Forecasting	Intermediate Output	✓ ✓	✓ ✗	✗ ✗	Concat.; Self-attention Offline Synthesis (MLP)	S-Bert	2024	No
	TimeCAP [42]	TS, Text	General	Classification	Input Intermediate Output	✗ ✓ ✓	✗ ✓ ✗	✓ ✗ ✗	LLM Generation Concat.; Self-attention, Retrieval Addition	Bert, GPT-4	2024	No
	TimeXL [31]	TS, Text	General	Classification Forecasting	Intermediate Output	✓ ✓	✓ ✗	✗ ✗	Concat., Prompt; LLM Reasoning Addition	Bert, S-Bert GPT-4o	2025	No
	Hybrid-MMF [37]	TS, Text	General	Forecasting	Intermediate	✓	✗	✗	Concat	GPT-4o	2024	Yes[4]
	Time-LLM [10]	TS, Text	General	Forecasting	Input Intermediate	✗ ✓	✗ ✓	✓ ✗	Meta-description Concat.; Self-attention	LLaMA, GPT-2	2024	Yes[5]
	Time-VLM [98]	TS, Text, Image	General	Forecasting	Input Intermediate	✗ ✓	✗ ✓	✓ ✗	Feat. Imaging, Meta-description Addition; Gating, Cross-attention	ViLT, CLIP BLIP-2	2025	No
	Unitime [35]	TS, Text	General	Forecasting	Input Intermediate	✗ ✓	✗ ✓	✓ ✗	Meta-description Concat.; Self-attention	GPT-2	2024	Yes[6]
	TESSA [30]	TS, Text	General	Annotation	Intermediate	✓	✓	✓	Prompt; RL; LLM Generation	GPT-4o	2024	No
	InstructTime [12]	TS, Text	General	Classification	Intermediate	✓	✓	✗	Concat.; Self-attention	GPT-2	2023	Yes[7]
	MATMCD [68]	TS, Text, Graph	General	Causal Discovery	Intermediate	✓	✓	✓	Prompt; LLM Reasoning; Supervision	Multiple	2025	No
	STG-LLM [54]	ST, Text	General	Forecasting	Intermediate	✓	✓	✗	Concat.; Self-attention	GPT-2	2024	No
	TableTime [72]	TS, Text	General	Classification	Input	✓	✗	✓	Prompt; Reformulate	Multiple	2024	Yes[8]
	ContextFormer [6]	TS, Tabular	General	Forecasting	Intermediate	✓	✓	✗	Addition; Cross-attention	No	2025	No
	Time-MQA [38]	TS, Text	General	Multiple	Input	✓	✗	✗	Prompt	Multiple	2025	Yes[9]
金融及医疗领域	MAN-SF [67]	TS, Text, Graph	Finance	Classification	Intermediate	✓	✓	✗	Bilinear; Graph Convolution	USE	2020	No
	Ramford et al. [3]	TS, Text TS, Image	Finance	Retrieval	Intermediate Output	✗ ✗	✓ ✗	✗ ✓	Supervision	S-Bert	2024	No
	Chen et al. [11]	TS, Text, Graph	Finance	Classification	Intermediate	✗ ✓	✗ ✓	✓ ✗	LLM Generation Concat.; Graph Convolution	ChatGPT	2023	No
	Xie et al. [81]	TS, Text	Finance	Classification	Input	✓	✗	✗	Prompt	ChatGPT	2023	No
	Yu et al. [89]	TS, Text	Finance	Forecasting	Input	✓	✗	✗	Prompt	GPT-4, Open LLaMA	2023	No
	MedTsLLM [5]	TS, Text, Tabular	Healthcare	Multiple	Intermediate	✓	✓	✗	Concat.; Self-attention	Llama2	2024	Yes[10]
	BeepLLM [95]	TS (Audio), Text	Healthcare	Classification	Intermediate	✓	✓	✗	Addition, Self-attention	OpenBioLLM-8B	2024	No
	METS [45]	TS, Text	Healthcare	Classification	Output	✗	✓	✗	Contrastive	ClinicalBert	2023	No
	Wang et al. [75]	TS, Text	Healthcare	Classification	Intermediate	✗	✗	✓	Supervision	Bart, Bert, RoBerta	2021	No
	EEG2TEXT [52]	TS, Text	Healthcare	Generation	Output	✗	✗	✓	Self-supervision, Supervision	Bart	2024	No
	MEDHMP [74]	TS, Text	Healthcare	Classification	Intermediate	✓	✓	✗	Concat.; Self-attention, Contrastive	ClinicalTS	2023	Yes[11]
	Dennabi et al. [16]	TS, Text	Healthcare	Classification	Intermediate	✓	✗	✗	Concat.	Bio+Clinical Bert	2021	Yes[12]
	Niu et al. [60]	TS, Text	Healthcare	Classification	Intermediate	✓	✓	✗	Concat.; Cross-attention	BioBERT	2023	No
	Yang et al. [86]	TS, Text	Healthcare	Classification	Intermediate	✓	✓	✗	Concat., Addition; Gating	ClinicalBERT	2021	Yes[13]
	Liu et al. [56]	TS, Text	Healthcare	Classification Regression	Input	✓	✗	✗	Prompt	PaLM	2023	Yes[14]
其它领域	aTP-LLM [24]	ST, Text	Traffic	Forecasting	Input	✓	✗	✓	Prompt; Meta-description	Llama2-7B-chat	2024	Yes[15]
	UrbanGPT [48]	ST, Text	Traffic	Forecasting	Input	✓	✗	✓	Prompt; Meta-description	Vicuna-7B	2024	Yes[16]
	CityGPT [19]	ST, Text	Mobility	Multiple	Input	✓	✗	✗	Prompt	Multiple	2025	Yes[17]
	MULAN [97]	TS, Text, Graph	IoT	Causal Discovery	Intermediate	✓	✓	✓	Addition; Contrastive; Supervision	No	2024	No
	MEA [82]	TS, Image	IoT	Anomaly Detection	Intermediate	✓	✓	✗	Addition; Cross-attention, Gating	No	2023	No
	Ekambaram et al. [18]	TS, Image, Text	Retail	Forecasting	Intermediate	✓	✓	✗	Concat.; Self & Cross-attention	No	2020	Yes[18]
	Skondrei et al. [70]	TS, Image, Text	Retail	Forecasting	Intermediate	✓	✓	✗	Concat.; Cross-attention	No	2024	Yes[19]
	VIMTS [96]	ST, Image	Environment	Imputation	Intermediate	✓	✓	✗	Concat.; Supervision	No	2022	No
	LITE [44]	ST, Text, Image	Environment	Forecasting	Intermediate	✓	✓	✗	Concat.; Self-attention	LLaMA-2-7B	2024	Yes[20]
	AV-HuBERT [69]	TS (Audio), Image	Speech	Classification	Intermediate	✓	✓	✗	Concat.; Self-attention	HuBert	2022	Yes[21]
	SpeechGPT [90]	TS(Audio), Text	Speech	Generation	intermediate	✓	✓	✗	Concat.; Self-attention	LLaMA-13B	2023	Yes[22]
	LA-GCN [83]	ST, Text	Vision	Classification	Intermediate	✗	✓	✗	Supervision	Bert	2023	Yes[23]

IV Applications & Future Research Directions

IV 应用领域与未来方向

应用领域

- **医疗健康**
  - 整合多种数据源（如电子健康记录、音频、脑电图、心电图及其他可穿戴传感器数据）
  - 针对音频、脑电图、心电图等数据开发的定制多模态方法，用于呼吸健康分类、心脏信号分析等医疗任务
- **金融**
  - 将市场时间序列（股价、交易量等）与新闻文本或社交媒体信息融合，辅助趋势预测
  - 金融市场中，融合多模态信息可以更全面地评估市场情绪
- **交通与环境**
  - 结合地理空间数据、交通流量序列与天气新闻报道，用于更精准的出行预测与区域规划
  - 环境监测中，多模态信息的整合能够有效解决数据缺失问题，提升环境时空预测的准确性
- **其它领域**
  - 物联网：融合传感器时序与日志文本或图像监控实现工业设备故障诊断
  - 零售：新品销售预测（结合图像、文本描述）

未来方向

大模型与跨模态推理、整合多模态信息构建决策支持系统、提高多模态数据质量

总结

全面概述了多模态时间序列分析的现状

系统梳理了数据集和方法 (>40种)、提出了统一的跨模态交互框架 (融合、对齐、转换)、讨论了实际应用和未来方向

谢 谢！