

多模态知识学习

徐童

中国科学技术大学

报告
概要



问题背景



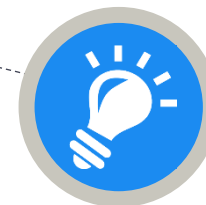
实体表征



关系学习



实体链接



未来展望

报告
概要



问题背景



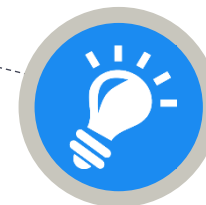
实体表征



关系学习



实体链接



未来展望

背景：我们身处在一个“多模态”的世界

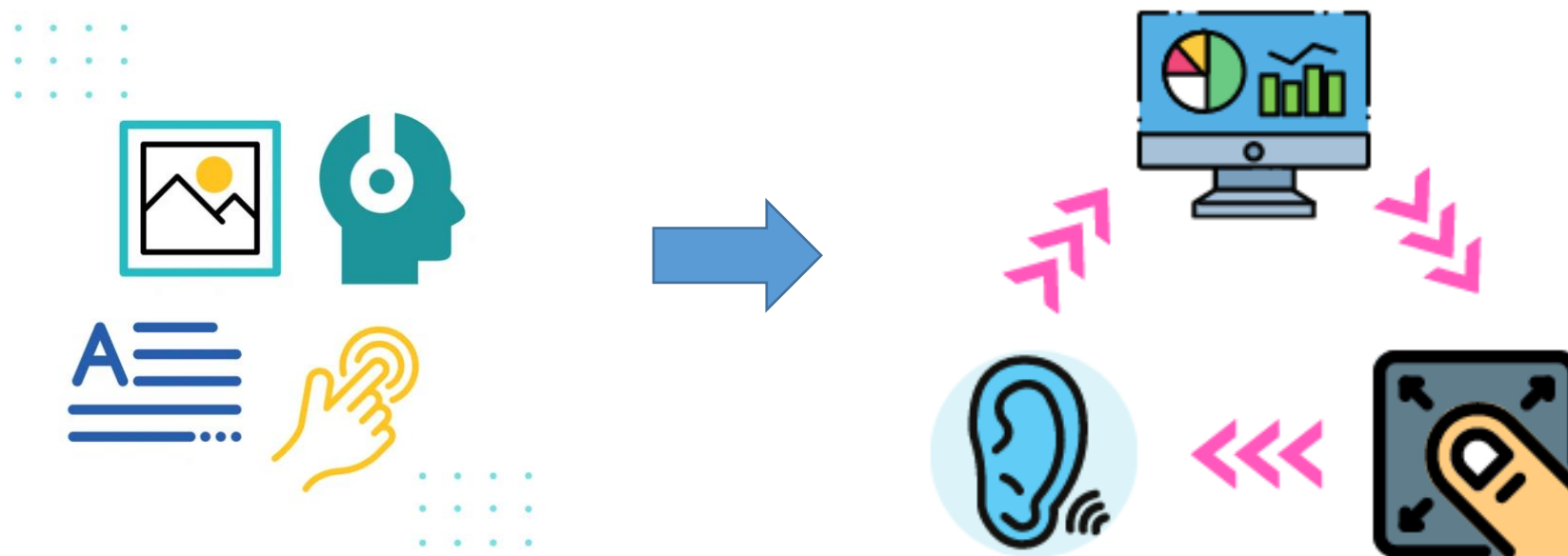
现阶段存在海量的多模态数据，人类对事物的认知也是基于多模态信息融合的



- 教育学研究表明，人类接受知识是通过多种方式的，其中视觉占83%，听觉占11%，嗅觉占3.5%
- 认知神经学研究表明，一类感官刺激可能会作用于其他感官通道，这种现象被称为“**联觉**”

背景：深度学习的发展为多模态联合学习奠定基础

随着海量多模态数据的积累和深度学习技术的发展，多模态学习正逐渐成为主流



- 机器翻译、图像识别、语义分割等常规任务可以通过多模态学习直接得到很大的效果提升。进一步将现有AI的认知模式组合会产生全新的应用场景，将推动人工智能从“感知智能”向“认知智能”转化

多模态学习案例

案例1：微软“小冰”机器人



- **微软小冰**不仅会作诗, 还会唱歌作曲、阅读朗诵、撰写新闻, 甚至可以生成虚拟人

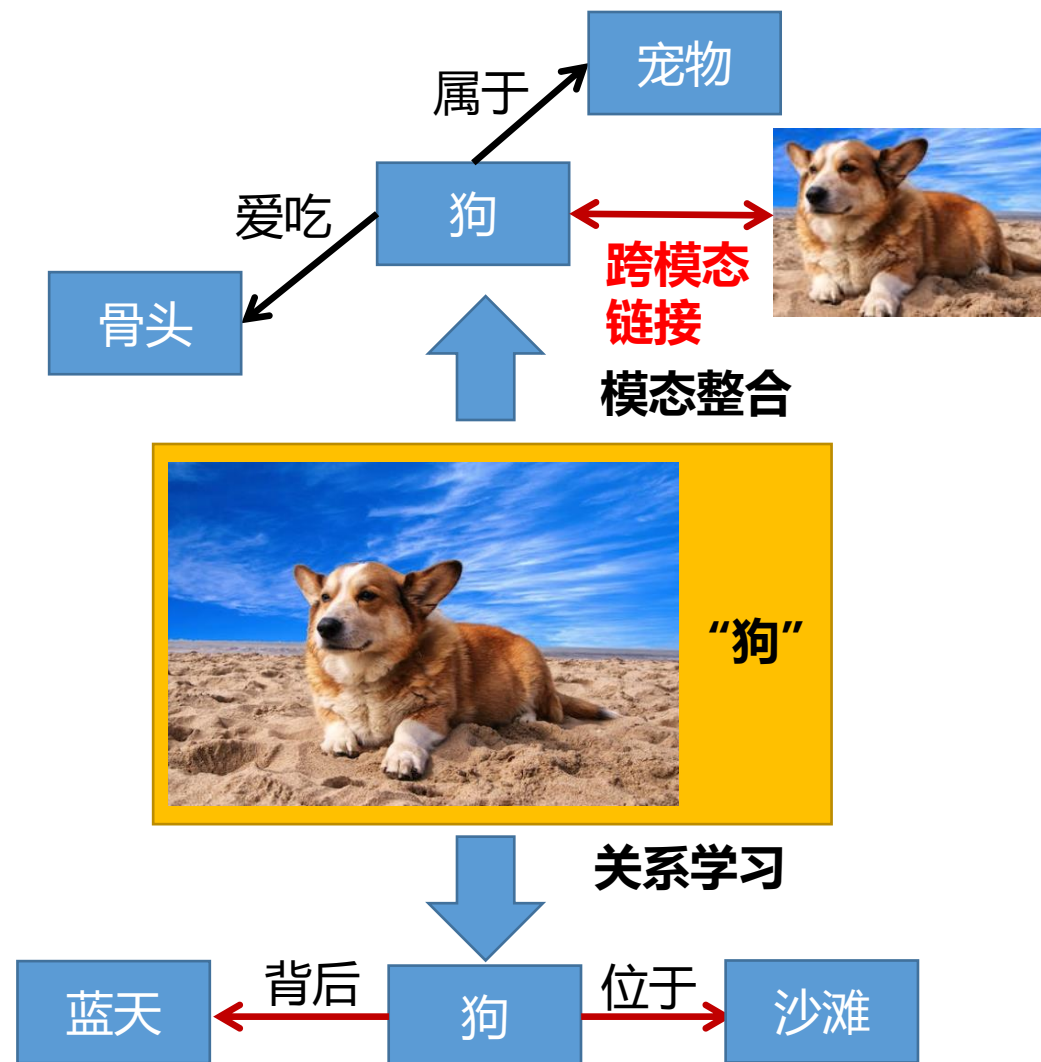
案例2：视频平台“只看TA”功能



- **优酷、爱奇艺**等推出“只看TA”功能, 可以在影视剧中筛选出特定演员的演出片段

需求：多模态数据，从“能用”到“好用”

- 面对多种模态的数据，有效获取并利用潜藏其中的**语义知识**，是实现“好用”的前提
 - 现有的知识大多是用符号文本的形式表示，迫切需要通过整合多模态数据，获取更加丰富完整的现实世界实体描述
 - 对于更加复杂的推理和理解类问题，必须从数据中获取更加准确全面的**结构化**知识



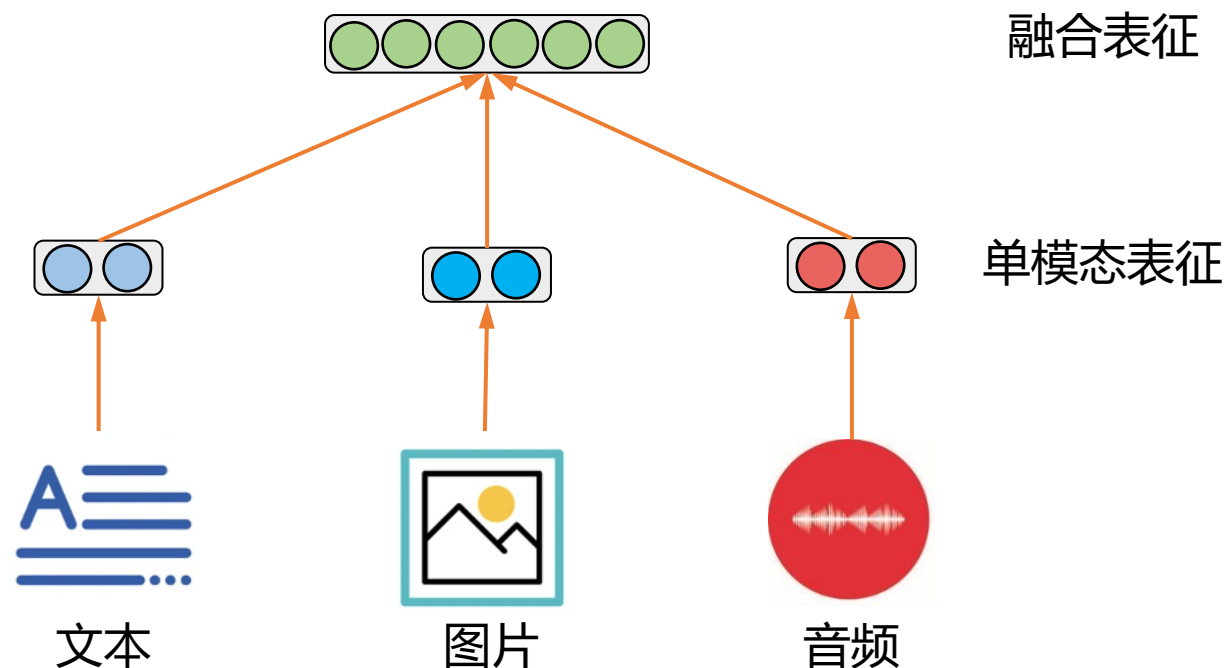
挑战：多模态知识学习的技术难点

(1) 机器从文本、图像、语音等多个模态获取信息，需要进行异构信息融合

➤ 多模态融合要求：

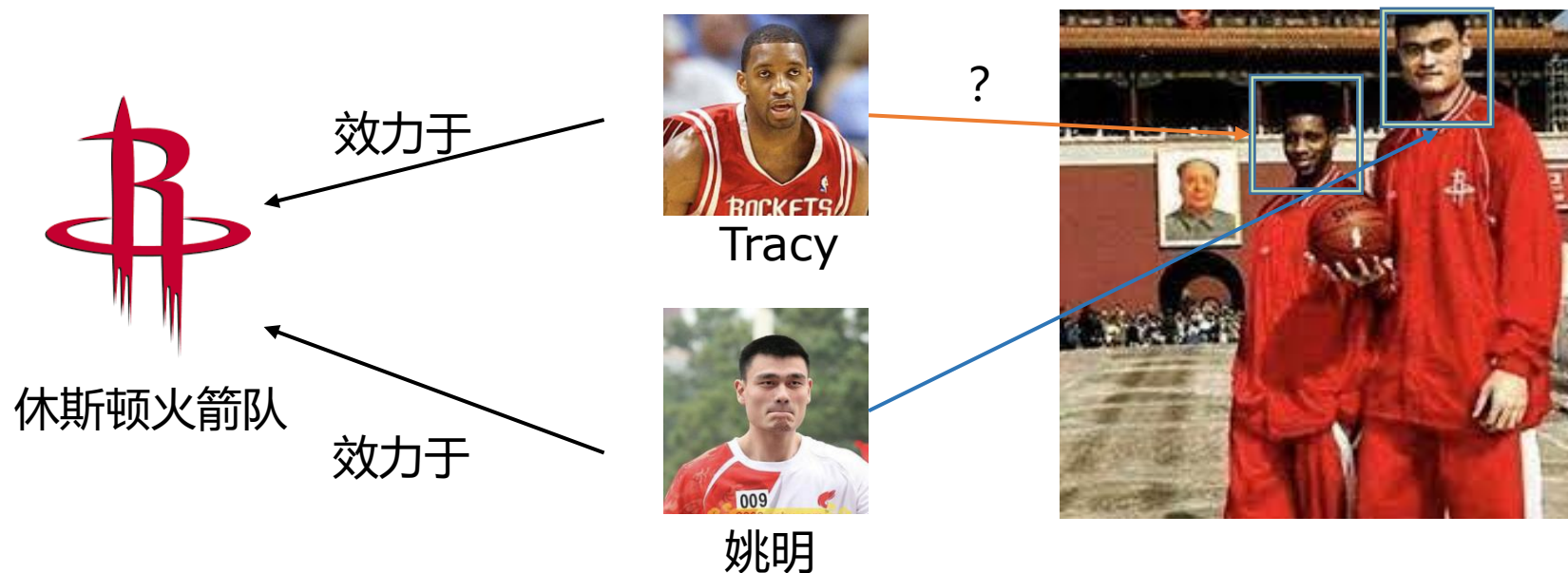
- 各个模态信息的多样性以及互补性
- 使各个模态的优点最大化
- 减少融合过程造成的信息损失

➤ 为实现多模态信息中实体单元的对齐，首先需要实现模态间的语义对齐



挑战：多模态知识学习的技术难点

(3) 即使获取了多模态知识，如何将信息链接到对应知识仍存在一定难度



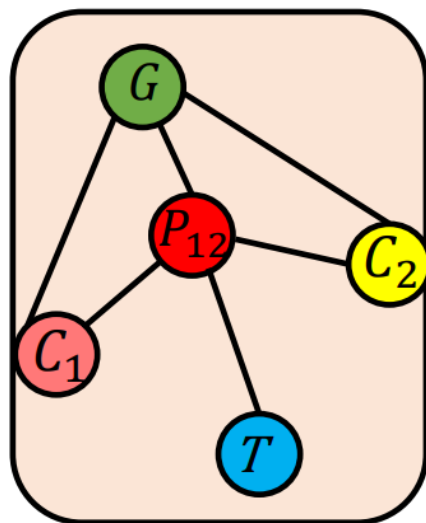
- 在多模态信息中，对于同一概念事物的表达，其形式要远比文本信息更为丰富，也更容易受到模态间或外部信息的干扰。同时，“隐语义”下的深层次线索也会干扰实体链接的效果。

任务：有效实现多模态知识的有效学习，并支撑下游应用

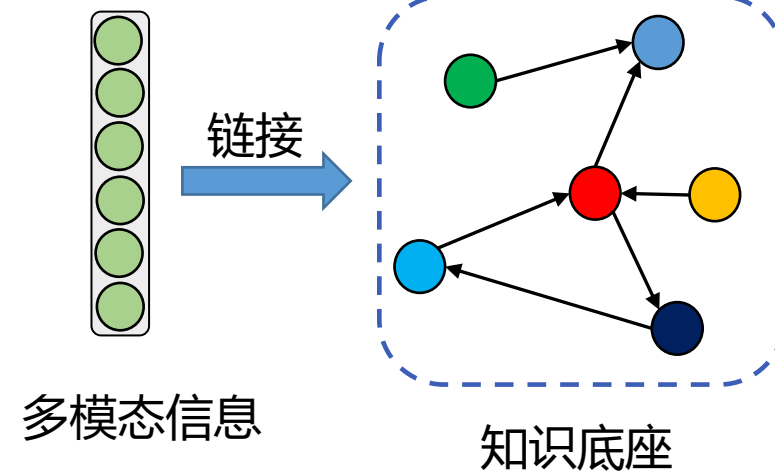
(a) 实体表征



(b) 关系学习



(c) 实体链接



报告
概要



问题背景



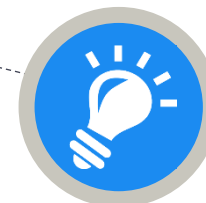
实体表征



关系学习



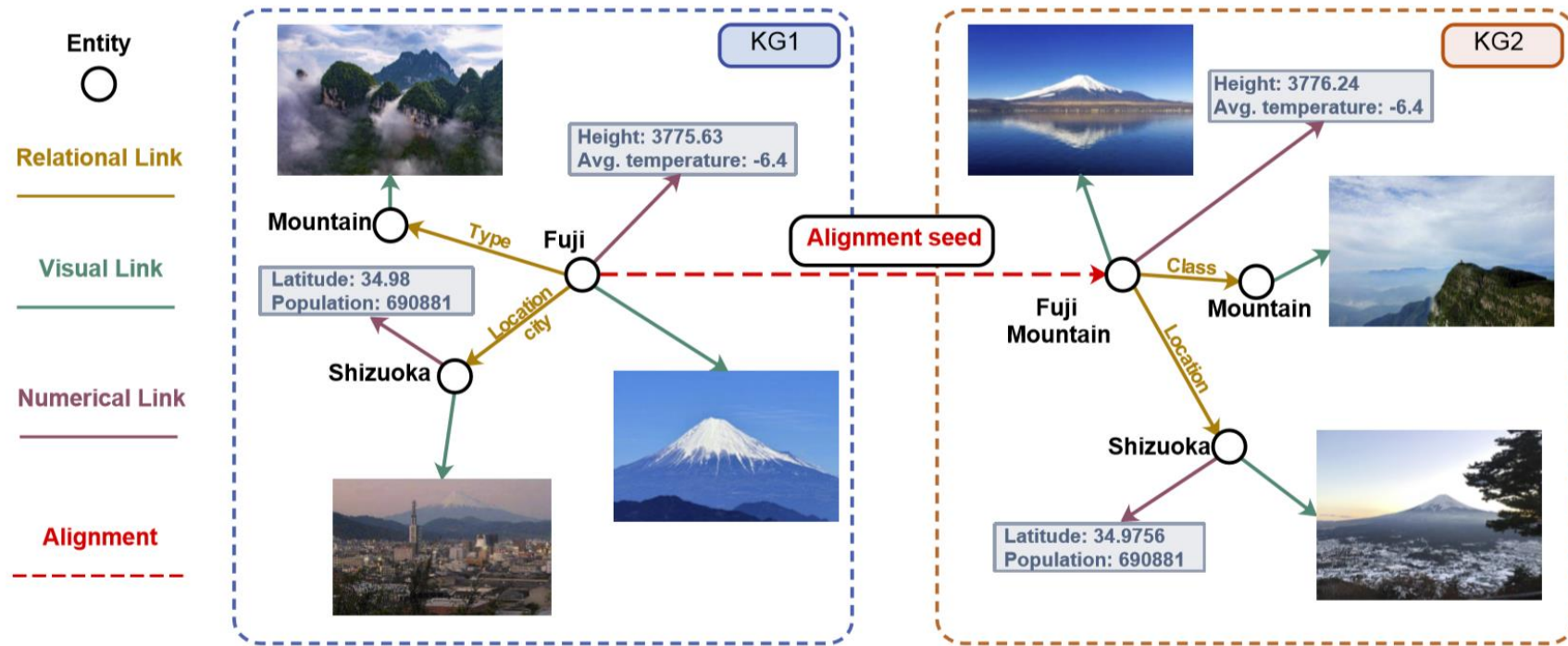
实体链接



未来展望

任务 (1) : 多模态实体表征

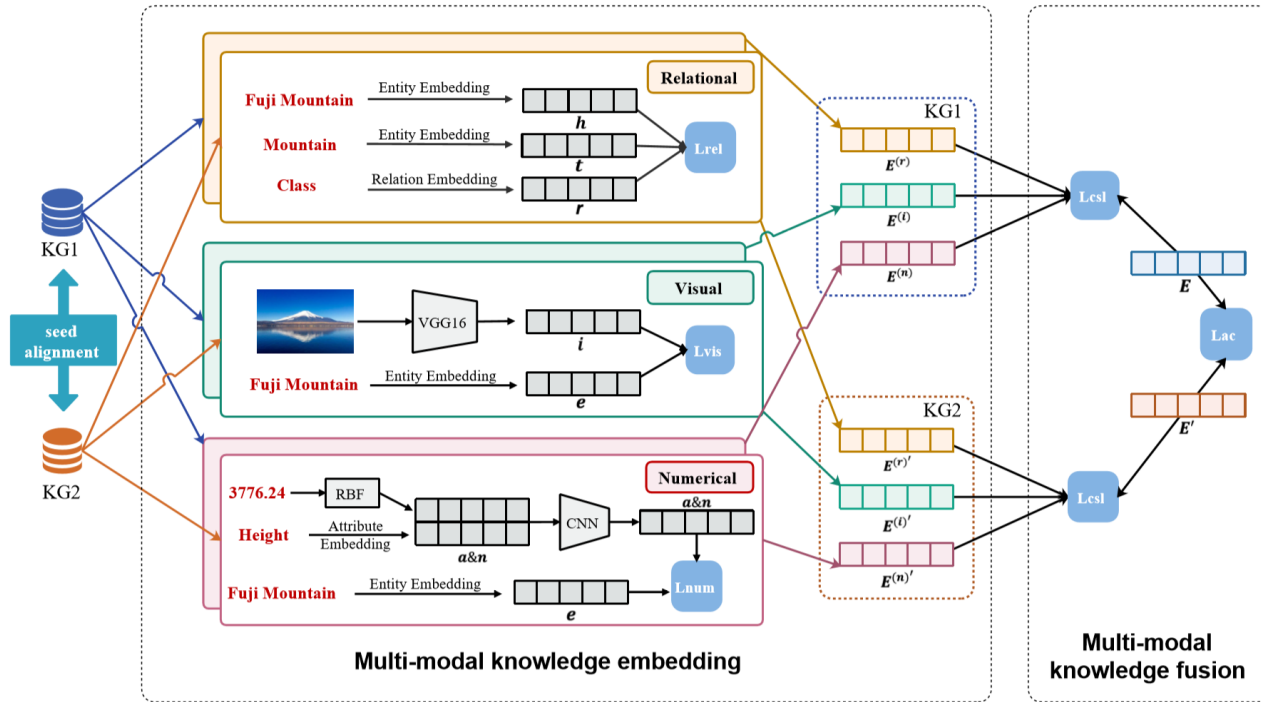
➤ 为什么需要实现多模态信息的融合表征与实体对齐?



➤ 在现实场景中，图像、文本、数值等多种模态存在语义互补现象，有助于我们完整理解知识。

任务 (1) : 多模态实体表征

- 基础思路：通过知识融合模块，使各个模态在语义空间层面实现统一

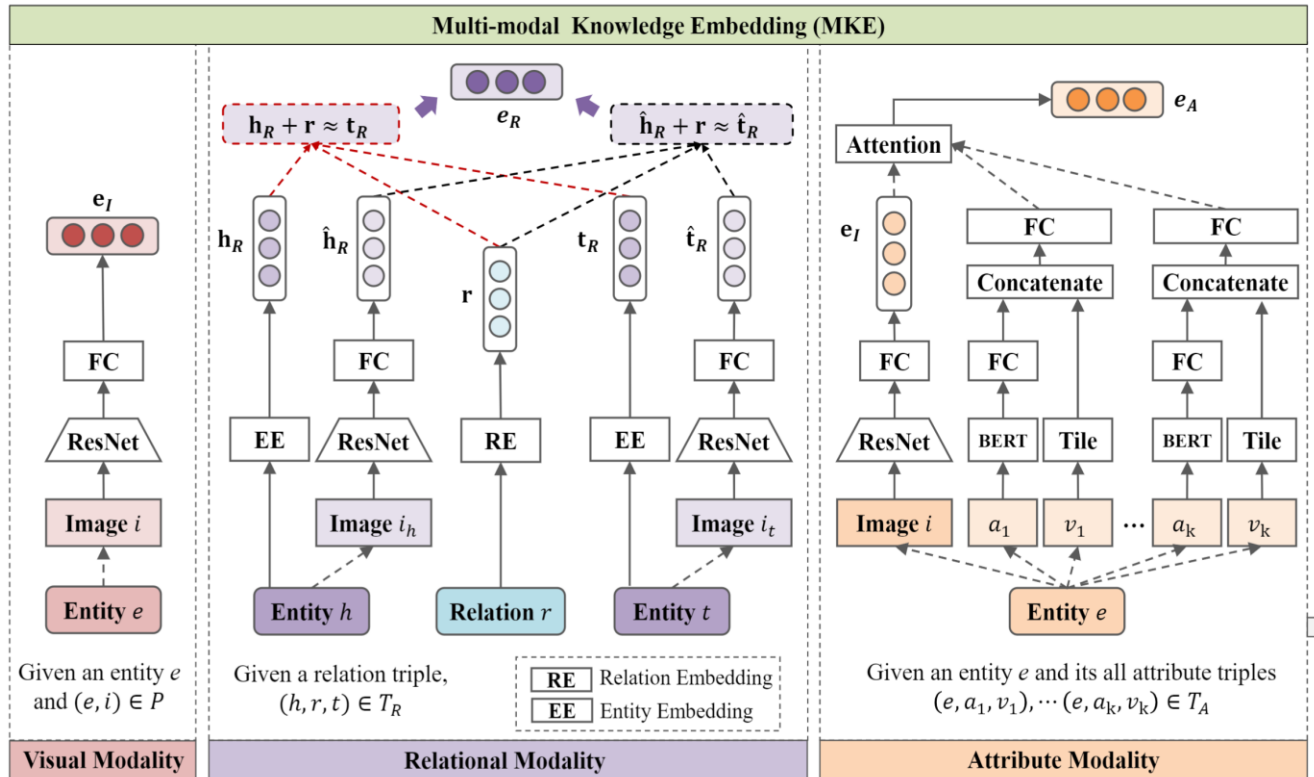


- 存在问题：各个模态语义表征空间相对独立，难以统一
- ✓ 首先，基于预训练模型分别进行不同模态知识的实体表示。
- ✓ 进而，以共同实体为中介，通过多模态知识融合模块，将不同表征约束至统一语义空间。

任务 (1) : 多模态实体表征

- 进阶思路: 强制融合并不能 “**扬长避短**”, 应捕捉各个模态中的高质量部分

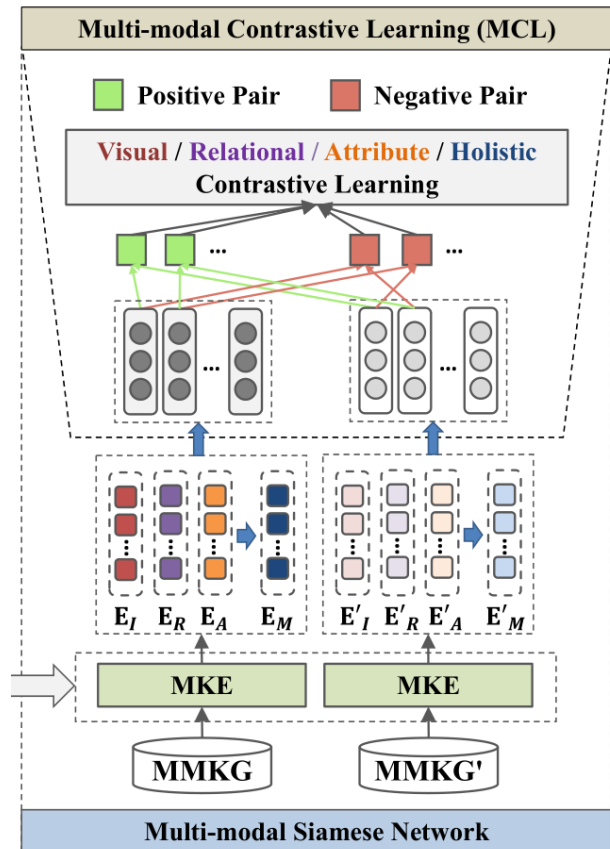
- 存在问题: 模态属性稀疏且表达简略, 笼统进行全局表征会造成干扰, 干扰融合进程
- 引入 **多模态增强机制**, 对模态内信息通过注意力捕获有价值的特征部分



任务 (1) : 多模态实体表征

- 进阶思路: 强制融合并不能 **“扬长避短”** , 应捕捉各个模态中的高质量部分

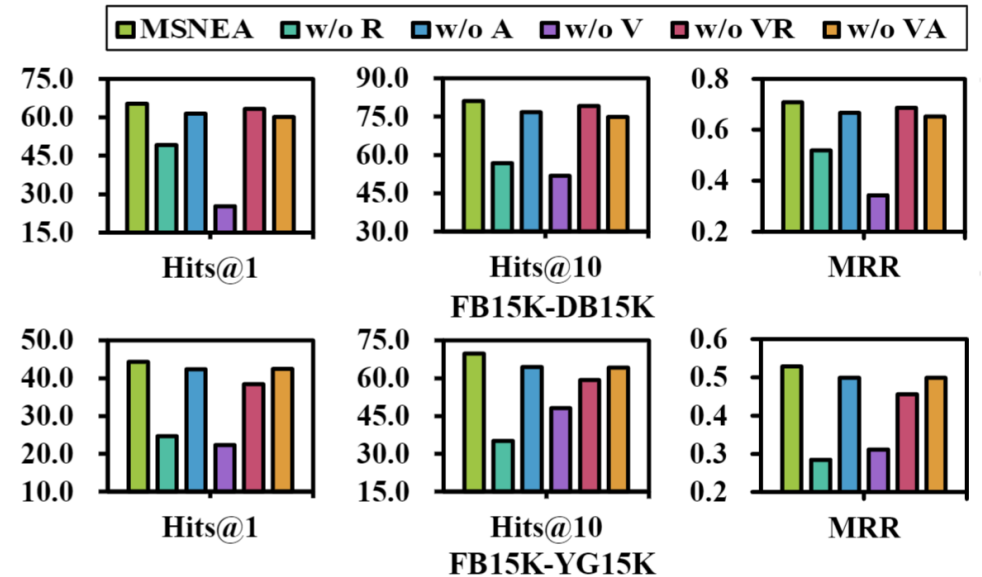
- 在模态间对齐过程中, 部分弱模态 (即高噪声模态) 会影响其他模态表征效果, 拖累整体精度
- ✓ 引入 多模态对比学习机制, 基于孪生网络, 在保持模态有效性的同时避免弱模态的过度影响, 实现多模态的增强融合



任务 (1) : 多模态实体表征

- 实验显示, 该模型在多模态实体对齐任务上达到了当时多模态知识图谱实体对齐模型的最佳性能。其中, **视觉信息**对模型性能有显著的影响。

Method	FB15K-DB15K					FB15K-YG15K				
	Hits@1	Hits@5	Hits@10	MR	MRR	Hits@1	Hits@5	Hits@10	MR	MRR
MTransE	0.359	1.414	2.492	1239.465	0.014	0.308	0.988	1.783	1183.251	0.011
IPTransE	3.985	11.226	17.277	387.512	0.086	3.079	9.505	14.443	522.235	0.070
GCN-Align	4.311	10.956	15.548	810.648	0.082	2.270	7.209	10.736	1109.845	0.053
BootEA	32.319	49.877	57.948	205.532	0.410	23.384	37.417	44.548	272.120	0.307
SEA	16.974	33.464	42.512	191.903	0.255	14.084	28.694	37.147	207.236	0.218
IMUSE	17.602	34.677	43.523	182.843	0.264	8.094	19.241	25.654	397.571	0.142
HyperKA	13.653	24.948	30.724	712.154	0.195	16.863	29.545	34.882	738.034	0.232
RAGA	6.101	11.482	14.965	1518.991	0.092	5.547	10.268	12.801	1337.759	0.082
RAC	20.285	35.954	43.205	453.313	0.281	15.066	28.103	34.494	501.795	0.216
PoE	12.0	-	25.6	-	0.167	10.9	-	24.1	-	0.154
MMEA	26.482	45.133	54.107	124.807	0.357	23.391	39.764	47.999	147.441	0.317
EVA	55.590	66.644	71.587	139.995	0.609	10.257	21.663	27.790	616.789	0.164
MSNEA	65.268	76.847	81.214	54.025	0.708	44.288	62.554	69.831	85.074	0.529



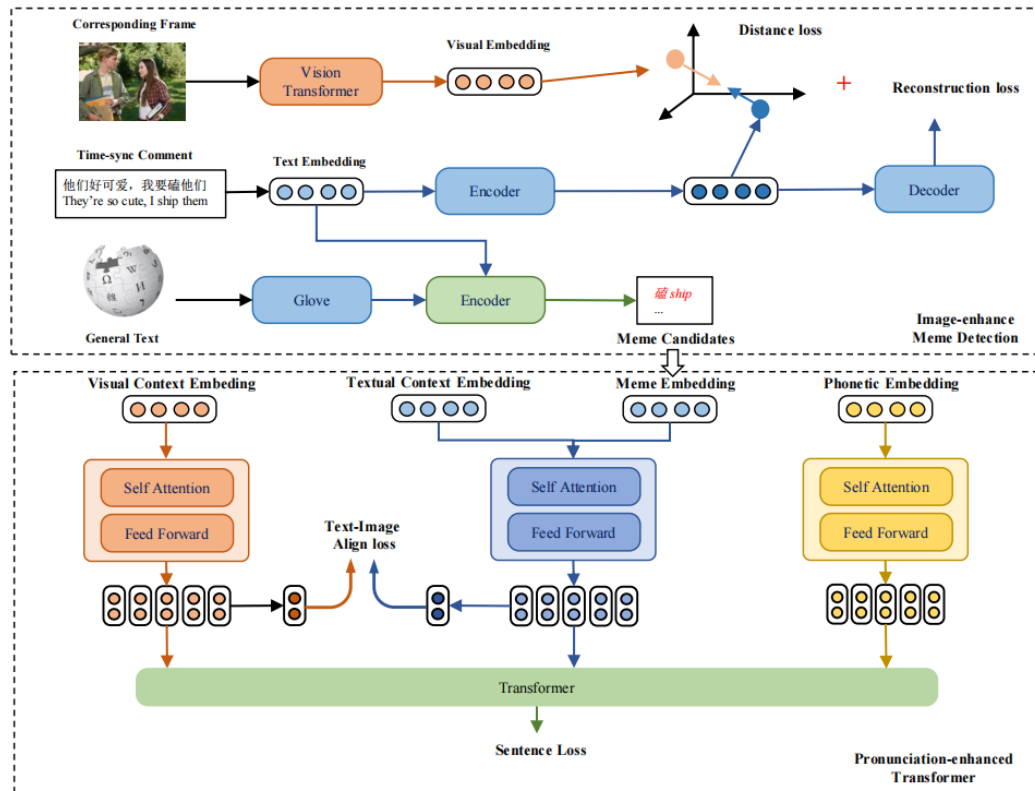
任务 (1) : 多模态实体表征

- 精益求精：跨模态表达同样存在歧义情况，不加区分将误导语义表征部分

➤ 存在问题：仅仅依靠单一模态信息，难以发现需要跨模态联想的深层次歧义元素

✓ 在融合跨模态信息作为约束的基础上，采用编解码结构重构原始文本信息，直至当前表达的真实语义表征

✓ 与规范化语义表达进行对比，检测并理解含有歧义的特殊表达

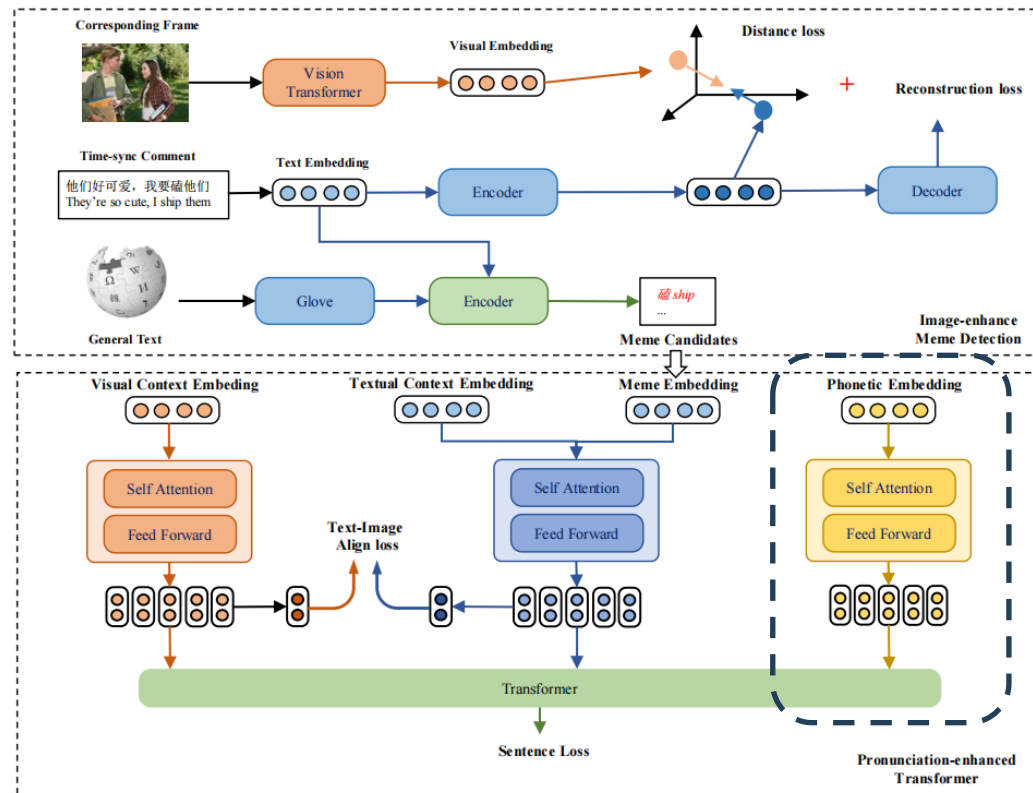


任务 (1) : 多模态实体表征

- 精益求精：跨模态表达同样存在歧义情况，不加区分将误导语义表征部分

➤ 存在问题：仅仅依靠单一模态信息，难以发现需要跨模态联想的深层次歧义元素

✓ 一个额外的小补丁：针对中文表达中“**同音词化用**”这一常见歧义类型，将**发音**作为额外模态引入融合



任务 (1) : 多模态实体表征

- 实验显示, 该模型不仅可以有效发现多模态语义表达中的歧义情况, 还可以为这些特殊表达生成对应的规范化表达“翻译”, 以实现更准确的多模态表征学习与语义理解

A/V(R)(S)	276/2174 (12.70%) (1000)			666/6677 (9.97%) (5000)			912/10839 (8.45%) (10000)		
Model	R@500	R@1000	R@1500	R@1600	R@3200	R@4800	R@2500	R@5000	R@7500
IMD-GRU	25.18	48.46	71.66	23.33	45.45	70.63	23.45	45.31	68.09
IMD-LSTM	21.98	46.58	69.08	22.51	46.84	69.73	21.71	45.89	67.76
IMD-Transformer	28.62	50.09	69.68	28.01	50.80	73.70	32.02	53.62	72.26
IMD-Transformer w/o distance loss	27.17	46.74	70.65	28.53	50.06	73.12	31.91	53.07	72.81
IMD-Transformer w/o reconstruction loss	25.36	48.91	68.12	29.28	49.70	71.17	26.10	49.89	70.18

Model	Q1-relevance	Q2-coherence
pointer generator	1.68	2.30
S2S-IC	2.64	3.06
S2S-IC2	3.07	3.46
BART	1.52	2.28
ProTE	3.28	3.90
w/o phonetic	3.24	3.75
w/o image	3.17	3.81
w/o alignment loss	2.78	3.49
average kappa	35.58%	32.60%

报告
概要



问题背景



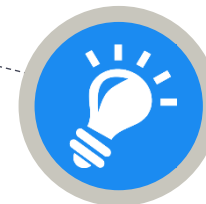
实体表征



关系学习



实体链接



未来展望

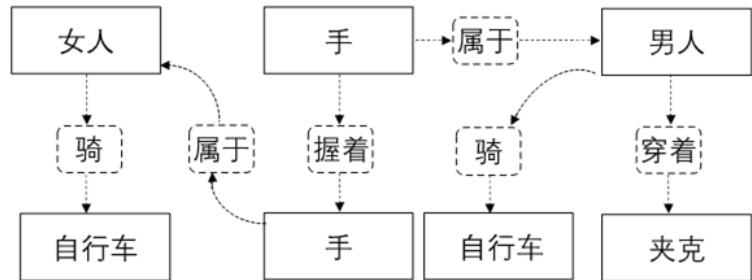
任务（2）：多模态关系学习

➤ 为什么需要实现多模态信息的结构化表达？

- 采用**图结构**规范表达目标及其关联
 - 节点表示实体目标，边表示实体间关系
- 基于结构化描述，机器可以通过推理解决复杂的描述/问答任务
 - 结构化描述的出现，不仅统一了各个模态信息的表示方式，更提升了跨模态信息的语义深度与可解释性



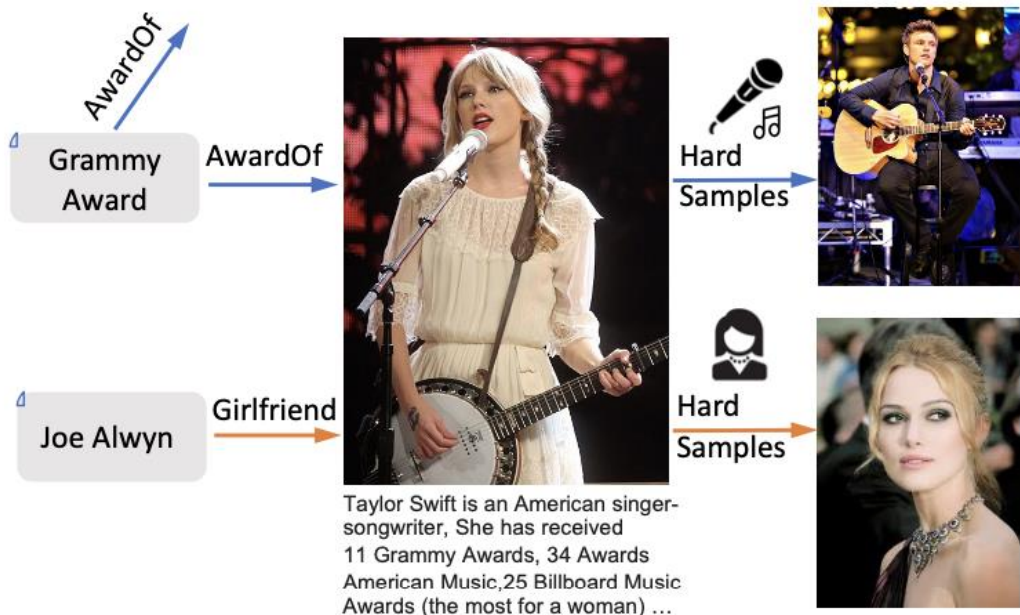
场景图生成



任务（2）：多模态关系学习

➤ 如何实现多模态信息的结构化表达？

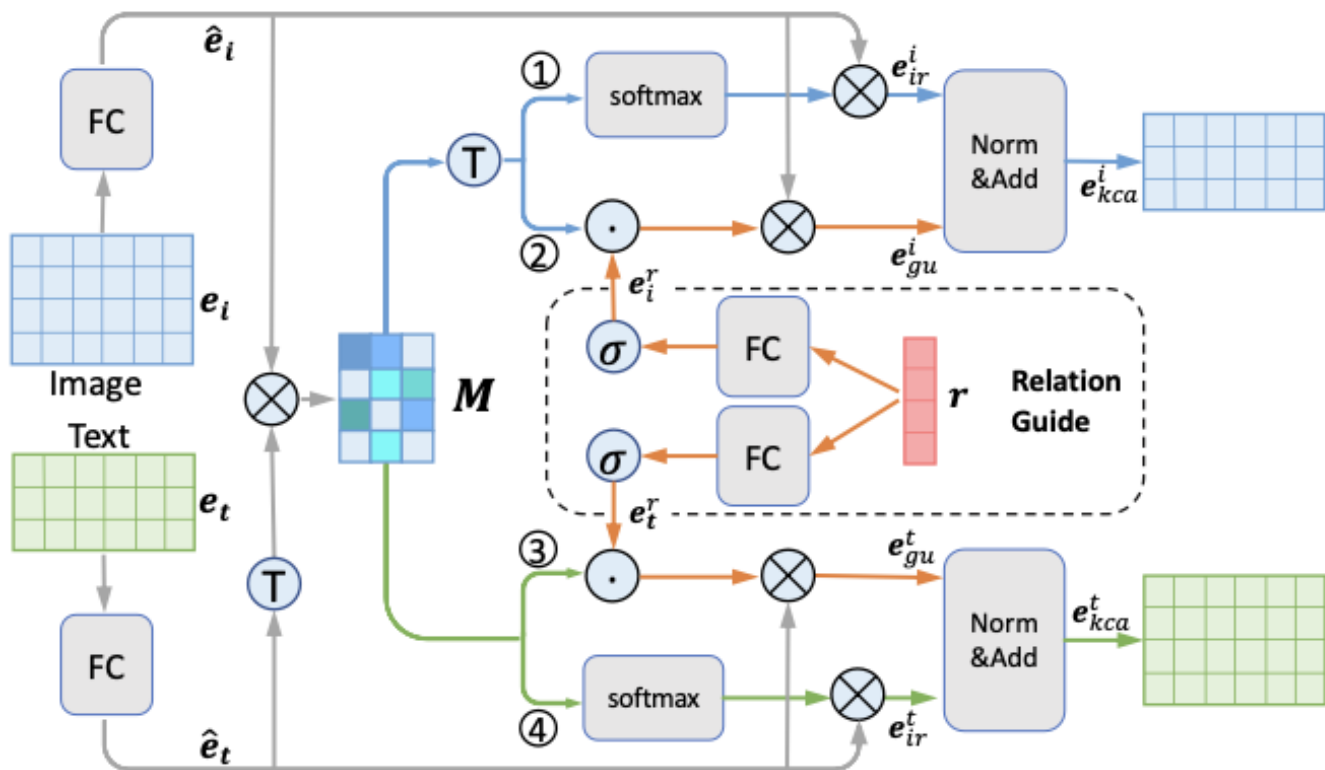
- 思路（1）：采用类似知识图谱关系补全的方法，补充完善目标间关系



- 考虑利用多模态语义信息（包括文本、图片）为KG中一个实体选取学习样本的场景。
- 需要解决的问题：目标间存在复杂的多类关系，同一个实体在不同关系对应下，模型训练的关注内容应该是不一样的。

任务（2）：多模态关系学习

➤ 如何实现多模态信息的结构化表达？



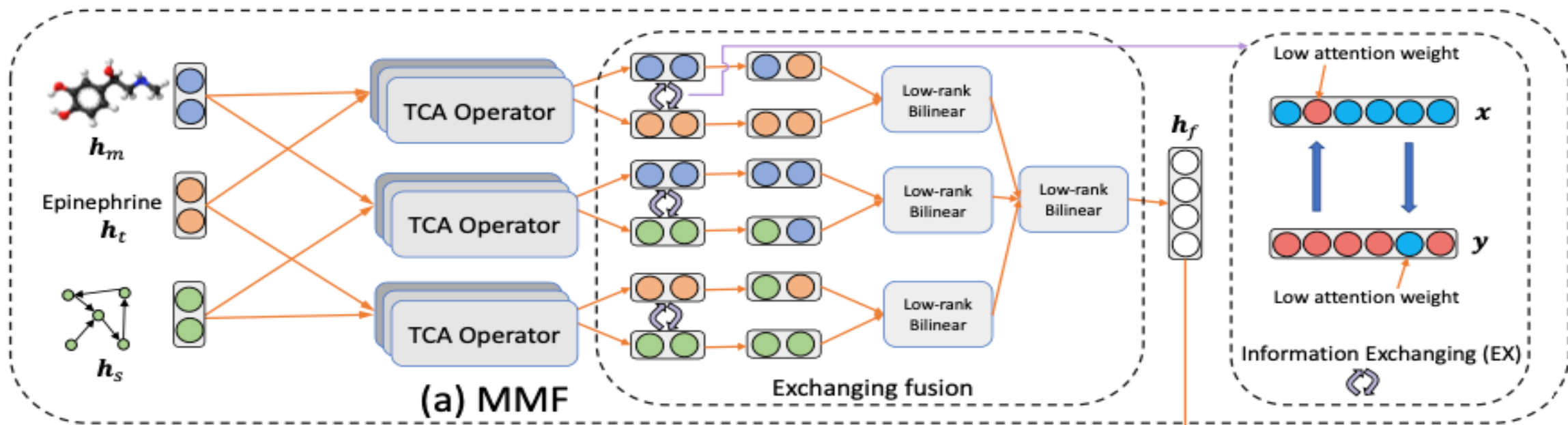
为此，我们提出了一种跨模态交叉的注意力机制，其中：

- ②、③分支用于利用知识引导学习多模态语义特征。
- ①、④分支用于衡量文本和图片模态之间的相互注意力。

任务 (2) : 多模态关系学习

➤ 如何实现多模态信息的结构化表达?

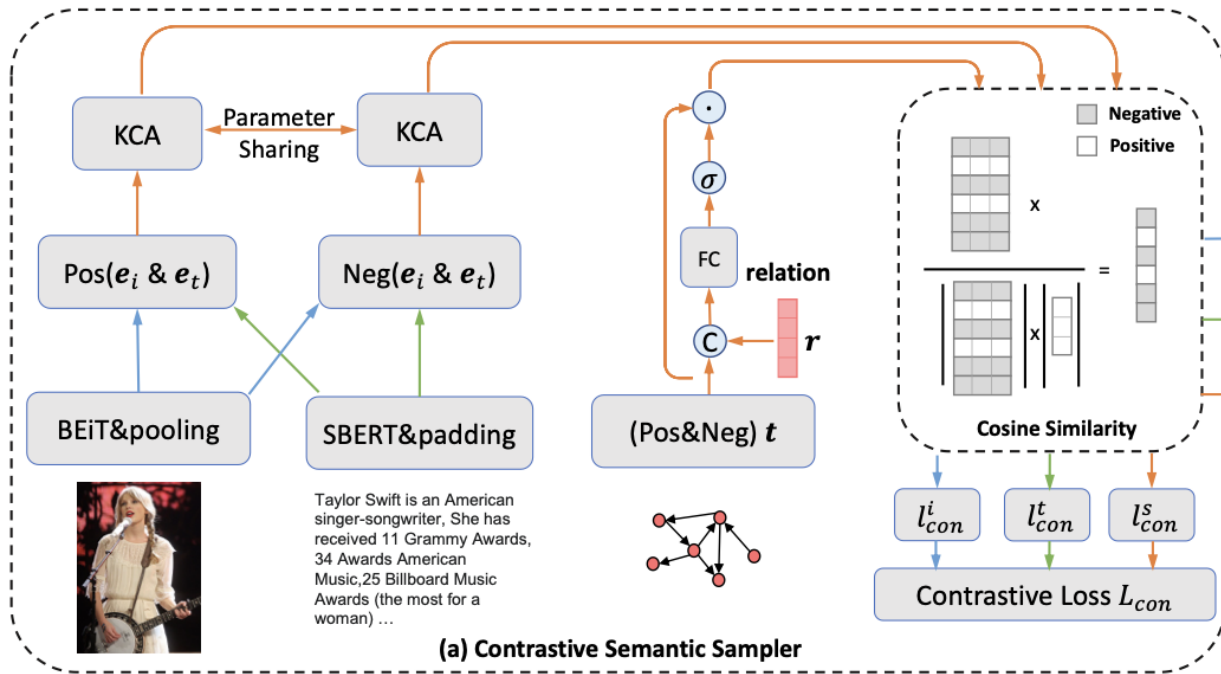
- 如果需要引入更多种类的模态信息? 没有问题, 采用两两模态对比学习可以有效解决
- 三元协同注意力 (TCA) 模块, 用于通过注意力赋权衡量信息重要性并弥合模态差距



任务 (2) : 多模态关系学习

➤ 如何实现多模态信息的结构化表达?

- 同时, 我们设计了对比语义采样器, 应对知识中普遍存在的1对多关系干扰

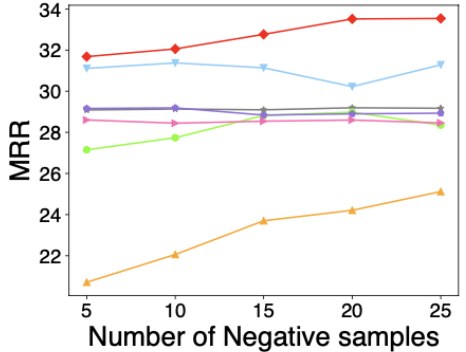
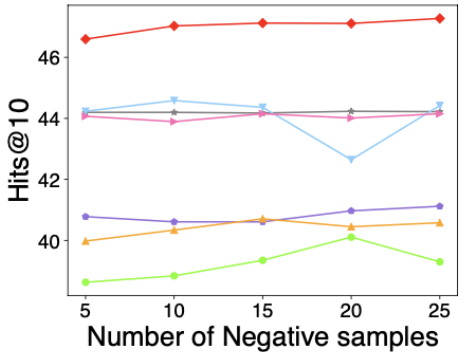
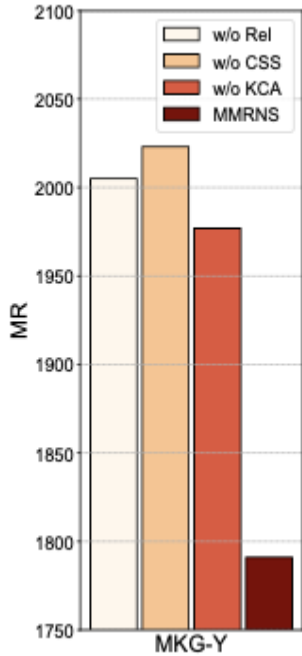
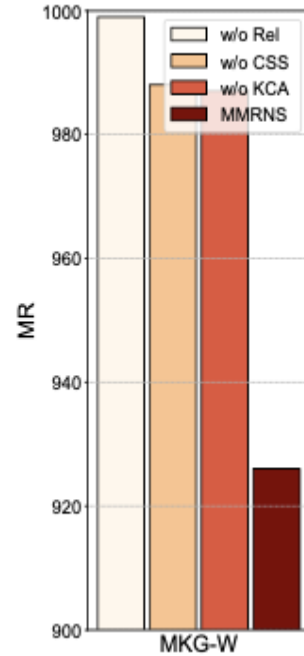
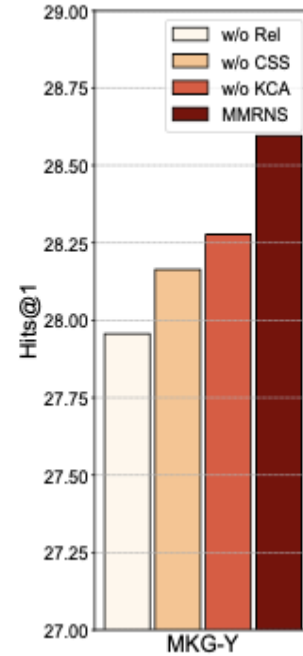
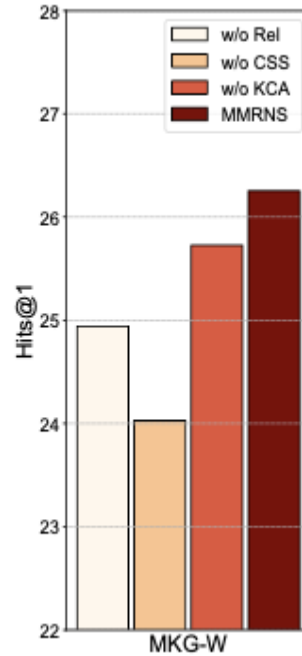
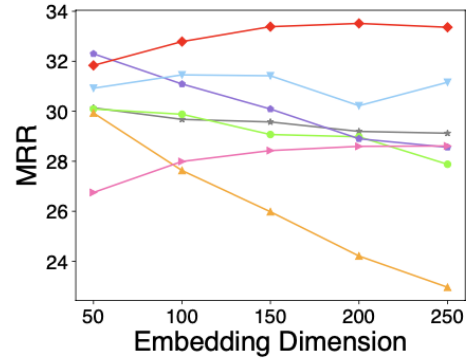
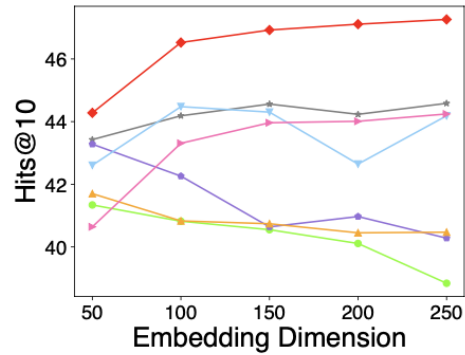


- 关系 AwardOf 和多个歌手 (实体) 有联系, 他们之间的语义特征应该是相近的。
- 构建对比语义采样器, 捕捉正样本之间语义特征的相似性, 和负样本之间的差异性。

任务 (2) : 多模态关系学习

- 参数实验发现: 在不同的 TransE 的模型参数下, 多模态实体关系补全都能有更好的性能提升。
- 消融实验发现: 去掉关系增强后的模型, 在多个指标下都出现下降, 体现了差异化处理的重要性。

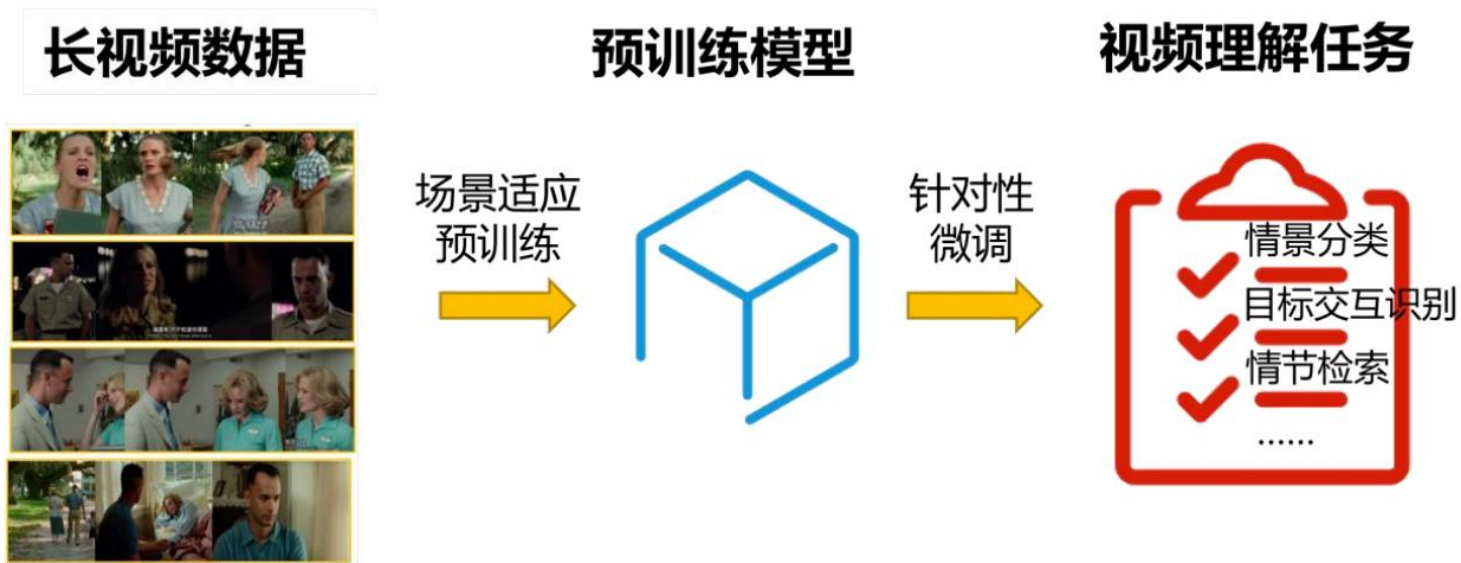
Legend: Uniform (grey line with dots), Bernoulli (green line with dots), Nscaching (purple line with dots), KBGAN (orange line with triangles), SANS (blue line with triangles), NS-KGE (pink line with triangles), MMRNS (red line with diamonds)



任务（2）：多模态关系学习

➤ 如何实现多模态信息的结构化表达？

- 思路（2）：借助大模型“见多识广”的特性，自主学习关系分类模型



任务（2）：多模态关系学习

➤ 如何实现多模态信息的结构化表达？

- 将目标关系识别任务转化成**视频问答任务**
 - ✓ 构建问题-答案查询模板，候选答案为所有可能的关系，其中真实关系作为正确答案
 - ✓ 输入预训练模型，预训练模型学习关系分类依据，获取答案作为关系分类

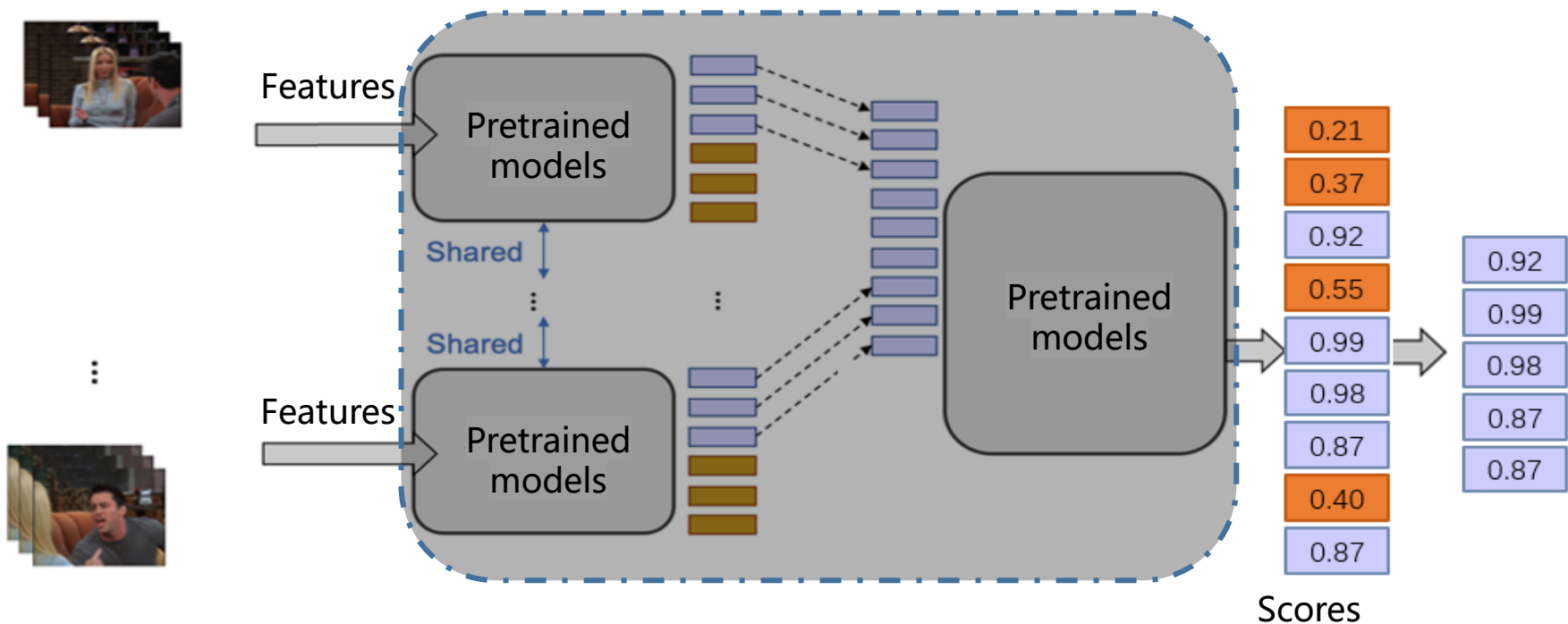


查询示例：{person1}和{person2}之间的**交互/社会关系**是什么？

任务（2）：多模态关系学习

➤ 如何实现多模态信息的结构化表达？

- 采用对比学习等技术，引导模型从上下文描述中提炼正确的分类依据，并过滤低质量信息
- 在获取实体关系之后，结合多个实体间的关系间接作用，通过图神经网络再进行二次优化



任务 (2) : 多模态关系学习

- 实验显示, 通过预训练模型, 可以在实体关系识别与补全的任务上获得有竞争力的结果
 - 该成果获得多媒体领域国际权威会议ACM MM 2022长视频理解竞赛 **两项赛道亚军**。

Task	Question Answer(6QA)		Building Graphs(68QA/116QA)		
	Acc	Hit@3	Acc	Hit@10	Hit@50
Movie-Level	26.12	60.18	2.95	25.05	85.08
Scene-Level	54.06	83.51	19.87	43.12	75.43

Max frame length	Negative sampling	R@1	R@5	R@10
140	sample	23.73	50.85	59.89
120	sample	26.55	50.28	66.10
120	full	24.86	49.15	63.84

Movie	S1	S2	S3	S4	S5	S6
Chained for life	15.00	71.00	50.00	33.00	80.00	0.00
Liberty kid	10.00	36.00	25.00	0.00	80.00	0.00
Like me	5.00	21.00	0.00	0.00	40.00	17.00
Little rock	0.00	0.00	0.00	50.00	60.00	50.00
Losing ground	5.00	19.00	25.00	75.00	80.00	33.00
Calloused hands	11.00	31.00	25.00	0.00	50.00	17.00
Mean	7.00	30.00	21.00	26.00	65.00	19.00

报告
概要



问题背景



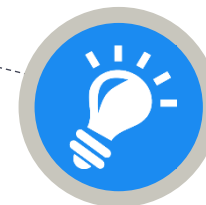
实体表征



关系学习



实体链接

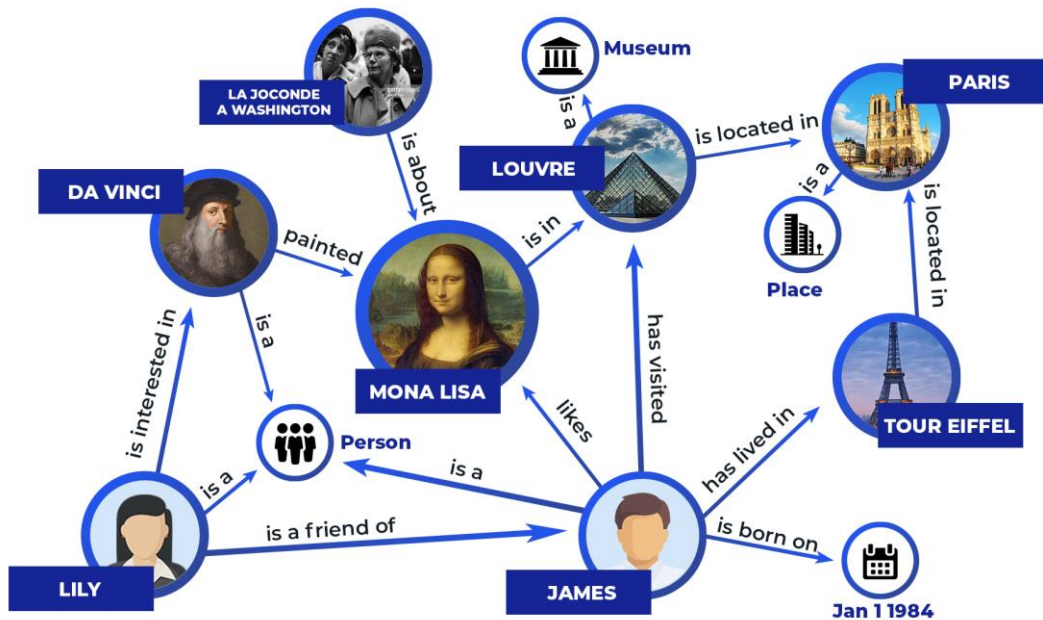


未来展望

任务 (3) : 多模态实体链接

➤ 为什么需要实现多模态信息的实体链接?

- 实现目标实体与多模态知识底座的有效链接, 是利用已抽取多模态知识的前提和基础。



任务 (3) : 多模态实体链接

➤ 多模态实体链接的挑战有哪些?

- 语义表达模态不统一：如何拉齐各个模态的表达方式差异（融合表征/结构化表示 ✓）
- 语义信息过于简略：短文本、缩写等过于精简、容易产生歧义的文本。
- 隐式语义难以挖掘：部分语义信息需要通过推理获得。

• 语义表达模态不统一

富士山是日本一座横跨静冈县和山梨县的活火山

.....




• 语义信息过于简略

.....
 作为特邀嘉宾，**莱昂纳多**也出席了活动。



莱昂纳多·威廉·迪卡普里奥
 演员、导演



莱昂纳多·纳西门托·德·阿劳若
 足球运动员

• 隐式语义难以挖掘



?

- 运动员
- 足球运动员
- ...
- 橄榄球运动员

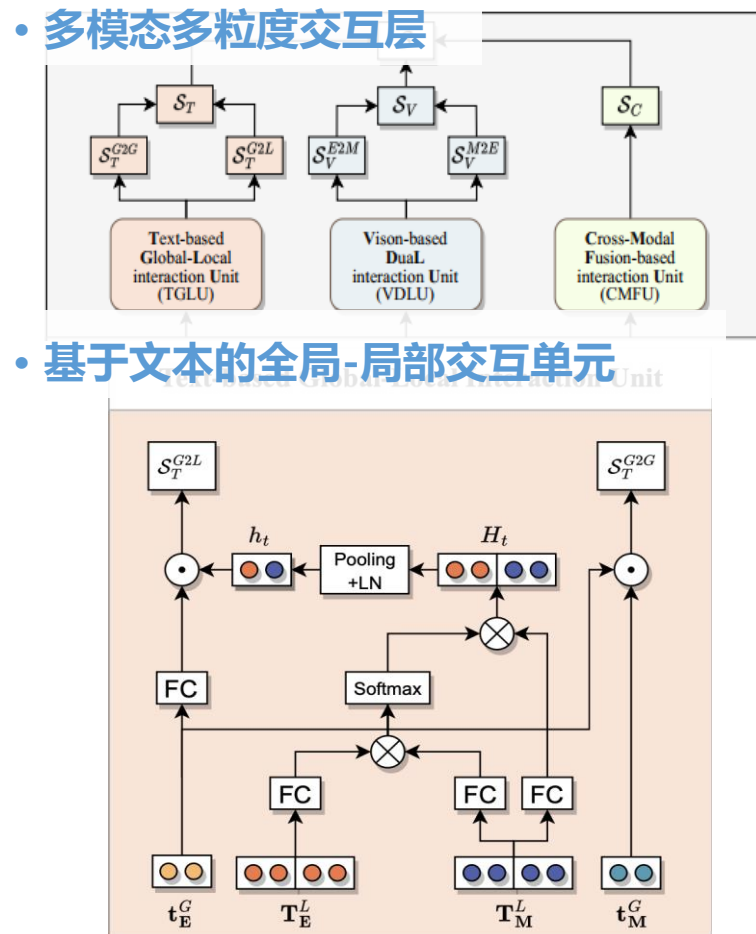
?

- 男性
- 女性

任务 (3) : 多模态实体链接

➤ 多模态实体链接的挑战有哪些?

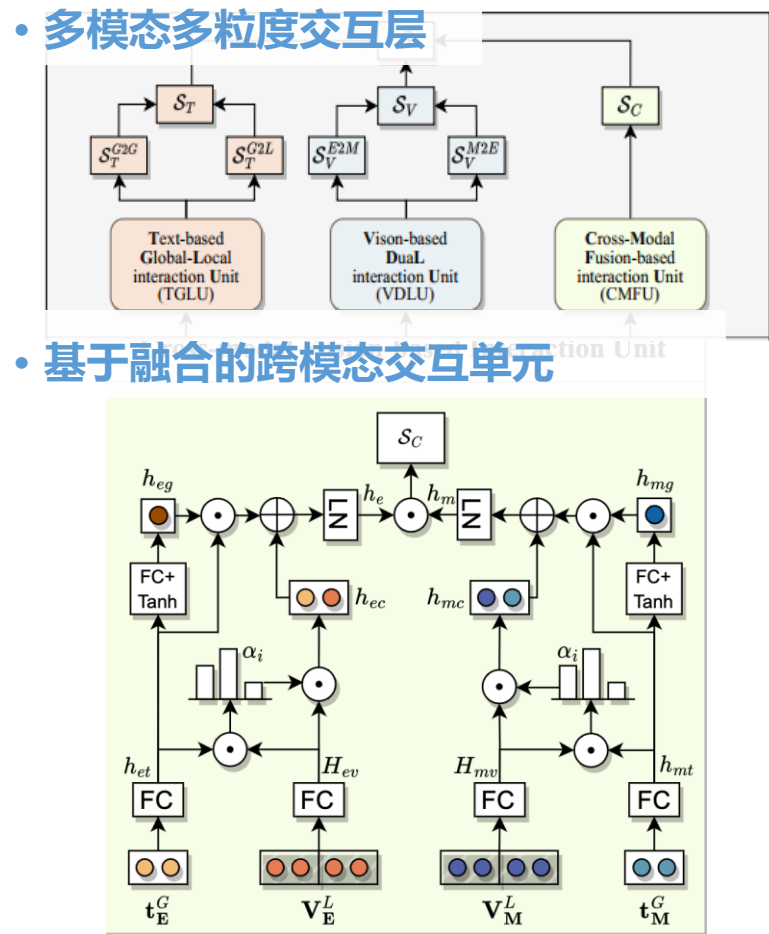
- 针对语义信息过于简略问题
 - ✓ 在多模态多粒度交互层中提出**全局-局部的交互单元**, 利用丰富知识弥补文本内容不足
 - ✓ 使用交叉注意力机制, 学习简略文本信息与其他模态的相关性, 从**全局**和**局部**两种角度评估其他模态信息对文本的补充强度。



任务 (3) : 多模态实体链接

➤ 多模态实体链接的挑战有哪些?

- 针对隐式语义难以挖掘问题
 - ✓ 在多模态多粒度交互层中, 提出一种基于融合的跨模态交互单元, 利用文本知识和图像之间的关联增强隐式语义
 - ✓ 使用门控机制, 控制不同来源的模态间的关联程度, 提升隐式语义相关性的表达



任务 (3) : 多模态实体链接

- 多个数据集上的实验证实了该方法在挖掘隐式语义和补充语义知识上的准确性
- 此外，低资源条件下实验和消融实验也分别验证了该模型的强大通用性

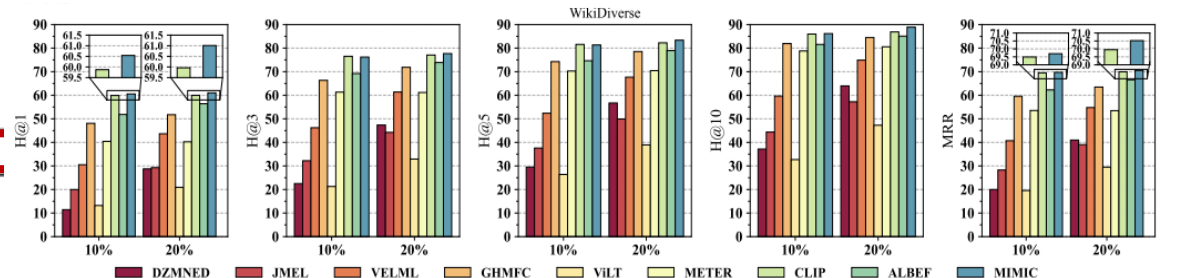
• 在三个数据集上与基线模型相比

Model	WikimEL					RichpediaMEL					WikiDiverse			
	H@1↑	H@3↑	H@5↑	MRR↑	MR↓	H@1↑	H@3↑	H@5↑	MRR↑	MR↓	H@1↑	H@3↑	H@5↑	MRR↑
BLINK [38]	74.66	86.63	90.57	81.72	51.48	58.47	81.51	88.09	71.39	178.57	57.14	78.04	85.32	69.15
BERT [9]	74.82	86.79	90.47	81.78	51.23	59.55	81.12	87.16	71.67	278.08	55.77	75.73	83.11	67.38
RoBERTa [23]	73.75	85.85	89.80	80.86	31.02	61.34	81.56	87.15	72.80	218.16	59.46	78.54	85.08	70.52
DZMNED [26]	78.82	90.02	92.62	84.97	152.58	68.16	82.94	87.33	76.63	313.85	56.90	75.34	81.41	67.59
JMEL [1]	64.65	79.99	84.34	73.39	285.14	48.82	66.77	73.99	60.06	470.90	37.38	54.23	61.00	48.19
VELML [43]	76.62	88.75	91.96	83.42	102.72	67.71	84.57	89.17	77.19	332.85	54.56	74.43	81.15	66.13
GHMFC [35]	76.55	88.40	92.01	83.36	54.75	72.92	86.85	90.60	80.76	214.64	60.27	79.40	84.74	70.99
CLIP [29]	83.23	92.10	94.51	88.23	17.60	67.78	85.22	90.04	77.57	107.16	61.21	79.63	85.18	71.69
ViLT [18]	72.64	84.51	87.86	79.46	220.76	45.85	62.96	69.80	56.63	675.93	34.39	51.07	57.83	45.22
ALBEF [21]	78.64	88.93	91.75	84.56	47.95	65.17	82.84	88.28	75.29	122.30	60.59	75.59	81.30	69.93
METER [11]	72.46	84.41	88.17	79.49	111.90	63.96	82.24	87.08	74.15	376.42	53.14	70.93	77.59	63.71
MIMIC	87.98*	95.07*	96.37*	91.82*	11.02	81.02*	91.77*	94.38*	86.95*	55.11*	63.51*	81.04	86.43*	73.44*

• 消融实验

Model	WikimEL						RichpediaMEL					
	H@1↑	H@3↑	H@5↑	H@10↑	H@20↑	MRR↑	H@1↑	H@3↑	H@5↑	H@10↑	H@20↑	MRR↑
MIMIC	87.98	95.07	96.37	97.80	98.73	91.82	81.02	91.77	94.38	96.69	98.04	86.95
w/o \mathcal{L}_T	86.13	93.69	95.74	97.66	98.57	90.42	72.82	89.05	93.12	96.15	97.61	81.61
w/o \mathcal{L}_V	86.71	94.43	96.25	98.01	98.80	90.94	78.72	90.23	93.66	96.04	97.61	85.15
w/o \mathcal{L}_C	86.67	94.04	95.69	97.21	98.18	90.74	79.65	89.89	92.56	94.92	96.94	85.38
w/o TGLU + \mathcal{L}_T	85.03	92.36	94.35	95.94	97.27	89.18	74.48	85.37	88.71	92.00	94.02	80.74
w/o VDLU + \mathcal{L}_V	83.46	93.33	95.47	97.23	98.18	88.74	74.12	89.47	92.81	95.82	97.61	82.37
w/o CMFU + \mathcal{L}_C	84.60	92.90	94.82	96.42	97.35	89.14	76.98	88.29	91.30	94.22	96.15	83.39

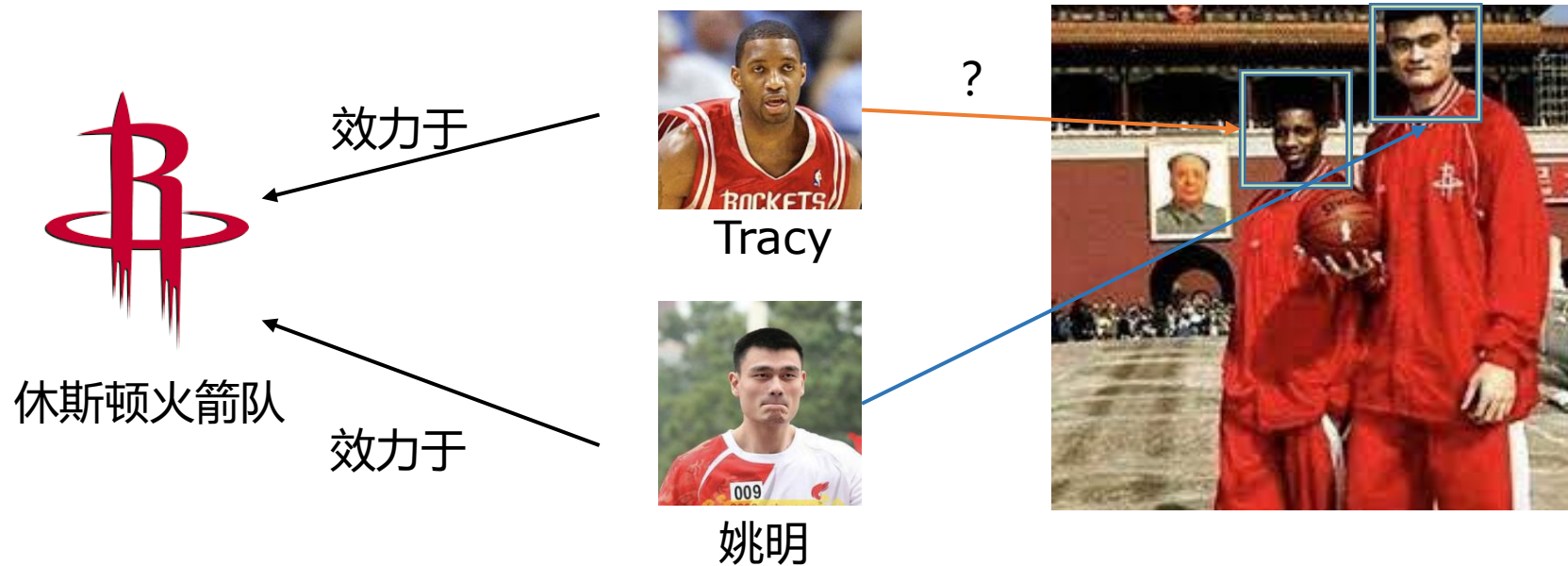
• 低资源实验



任务 (3) : 多模态实体链接

➤ 还有哪些信息有助于实现多模态信息的实体链接?

- 辅助实体链接判定的重要线索: 具有特定关联的、高置信度的共现实体



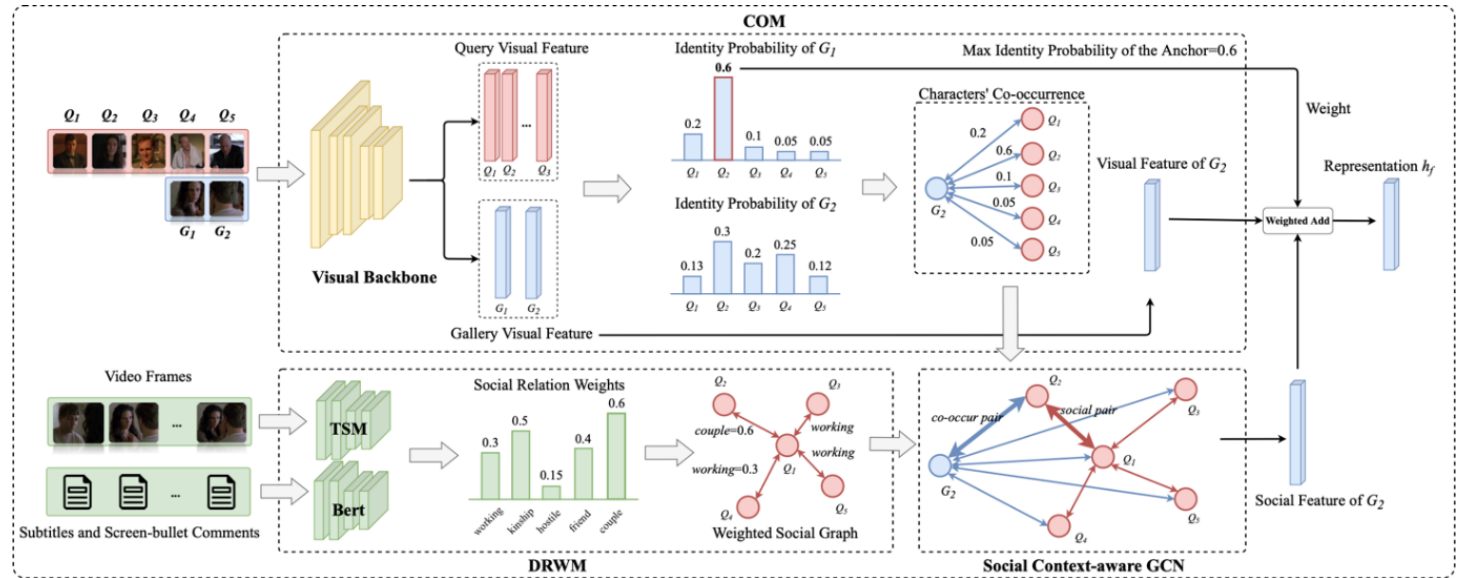
任务 (3) : 多模态实体链接

➤ 还有哪些信息有助于实现多模态信息的实体链接?

- 关联能否转化为共现关系，往往取决于所在的特定场景，需要首先识别当前场景下的共现概率

✓ **关系权重模块**：识别当前场景，并为不同关系生成共现概率

✓ **共现挖掘模块**：挖掘潜在目标对象，并建立加权共现关系图



任务 (3) : 多模态实体链接

- 真实世界数据集及消融实验证实，引入场景先验共现关系对于实体链接任务的有效提升

TABLE I

OVERALL PERFORMANCE ON TWO DATASETS. TRM MEANS TRANSFORMER. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Method	Backbone	Social-Bilibili			Social-MoviNet		
		mAP	mINP	R1	mAP	mINP	R1
HACNN [16]	CNN	47.8	29.4	66.0	30.4	20.7	55.0
MLFN [5]	CNN	55.7	34.6	72.0	30.6	20.6	40.0
ResNet [17]	CNN	61.2	35.8	82.0	42.1	22.2	65.0
ResNet-mid [4]	CNN	63.8	35.3	82.0	45.0	22.0	75.0
OSNet [1]	CNN	64.6	38.2	80.0	49.1	21.2	75.0
OSNet-AIN [20]	CNN	65.1	39.7	86.0	45.3	20.9	80.0
TP-ResNet [3]	CNN	59.5	36.2	78.0	44.8	21.7	75.0
NFormer [21]	TRM	63.8	36.3	80.0	37.8	20.7	65.0
ViT [22]	TRM	76.5	49.1	90.0	59.4	23.0	90.0
TransReID [7]	TRM	79.9	51.2	90.0	62.3	24.6	90.0
TP-ViT [3]	TRM	68.1	40.3	86.0	59.2	24.1	95.0
SCPS-L [3]	TRM	80.9	52.6	90.0	57.9	23.9	80.0
SCPS-P [3]	TRM	81.0	53.1	90.0	58.9	24.1	80.0
Ours	TRM	82.8	56.3	94.0	63.6	24.5	95.0

TABLE III

THE PERFORMANCE OF OUR METHOD WITH DIFFERENT MODAL INFORMATION ON SOCIAL-BILIBILI DATASET

Method	Visual	Textual	mAP	mINP	R1
w/o VT	-	-	79.9	51.2	90.0
w/o V	-	Bert	82.6	55.4	94.0
w/o T	TSM	-	82.7	56.2	94.0
all	TSM	Bert	82.8	56.3	94.0

TABLE IV

THE PERFORMANCE OF OUR METHOD WITH DIFFERENT BALANCE STRATEGIES ON SOCIAL-BILIBILI DATASET

Strategy	mAP	mINP	R1
Mean	82.4	55.4	94.0
Random	76.0	47.9	90.0
Adaptive(Ours)	82.8(+0.4)	56.3(+0.6)	94.0(+0)

报告
概要



问题背景



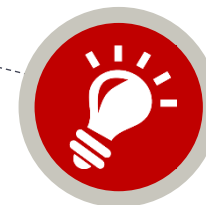
实体表征



关系学习



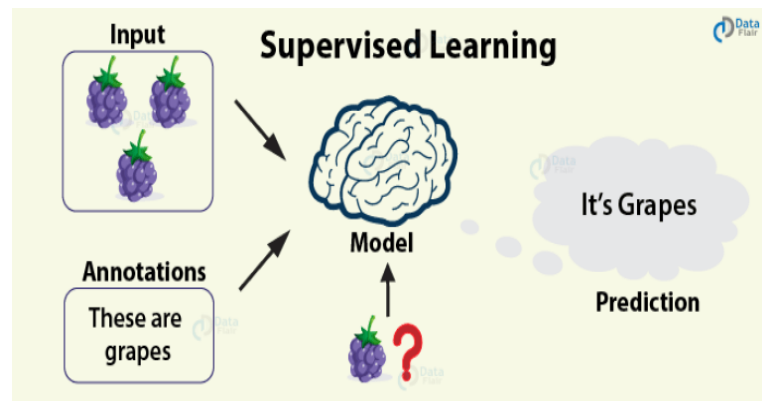
实体链接



未来展望

展望：如何更进一步发挥多模态知识学习的价值？

- 大模型时代来临，多模态知识学习的作用或更加凸显
 - 我们不仅需要丰富的联想，也需要严谨的推理，这有赖于多模态知识与大模型的进一步结合



- 更强的推理能力
- 更多的开放场景
- 联想与“涌现”
-

参考文献

1. Liyi Chen, Zhi Li, Yijun Wang, Tong Xu, Zhefeng Wang, Enhong Chen, MMEA: Entity Alignment for Multi-Modal Knowledge Graphs, In Proceedings of the 13th International Conference on Knowledge Science, Engineering and Management (KSEM'20), Hangzhou China, 2020, 134-147. (Best Paper Award)
2. Liyi Chen, Zhi Li, Tong Xu, Han Wu, Zhefeng Wang, Nicholas Yuan, Enhong Chen, Multi-modal Siamese Network for Entity Alignment, In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'22), Washington DC, USA, 2022, Accepted.
3. Shiwei Wu, Weidong He, Tong Xu, Hao Wang, Enhong Chen, Winning the CVPR'2022 AQTC Challenge: A Two-stage Function-centric Approach, arXiv preprint arXiv:2206.09597, 2022
4. Zheyong Xie, Weidong He, Tong Xu*, Shiwei Wu, Chen Zhu, Ping Yang, Enhong Chen, Comprehending the Gossips: Meme Explanation in Time-Sync Video Comment via Multimodal Cues , ACM Transactions on Asian and Low-Resource Language Information Processing , 2023, Accepted.
5. Penggang Qin, Jiarui Yu, Yan Gao, Derong Xu, Yunkai Chen, Shiwei Wu, Tong Xu*, Yanbin Hao, Enhong Chen, Unified QA-aware Knowledge Graph Generation Based on Multi-modal Modeling, In Proceedings of the 30th ACM International Conference on Multimedia (ACM MM'22), Lisbon, Portugal, 2022, Accepted.
6. Derong Xu, Tong Xu, Shiwei Wu, Jingbo Zhou, Enhong Chen, Relation-enhanced Negative Sampling for Multimodal Knowledge Graph Completion, In Proceedings of the 30th ACM International Conference on Multimedia (ACM MM'22), Lisbon, Portugal, 2022, Accepted.
7. Derong Xu, Jingbo Zhou, Tong Xu*, Yuan Xia, Ji Liu, Enhong Chen, Dejing Dou, Multimodal Biological Knowledge Graph Completion via Triple Co-attention Mechanism, In Proceedings of the IEEE International Conference on Data Engineering 2023 (ICDE'2023), Anaheim, California, USA, 2023, Accepted.
8. Pengfei Luo, Tong Xu*, Shiwei Wu, Chen Zhu, Linli Xu, Enhong Chen, Multi-Grained Multimodal Interaction Network for Entity Linking, In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'23), Long Beach, CA, USA, 2023, Accepted.
9. Wenjun Peng, Weidong He, Derong Xu, Tong Xu*, Chen Zhu, Enhong Chen, Social Context-aware GCN for Video Character Search via Scene-prior Enhancement, In Proceedings of the 2023 IEEE International Conference on Multimedia and Expo (ICME'2023), Brisbane, Australia, 2023, Accepted.

Thanks!

tongxu@ustc.edu.cn