



基于大模型的文本情感计算

周杰 青年研究员
计算机科学与技术学院



華東師範大學
EAST CHINA NORMAL
UNIVERSITY



華東師範大學
EAST CHINA NORMAL
UNIVERSITY

目录 | CONTENT

- 情感计算简介
- 预训练情感计算
- 生成式情感计算
- 情感计算发展趋势

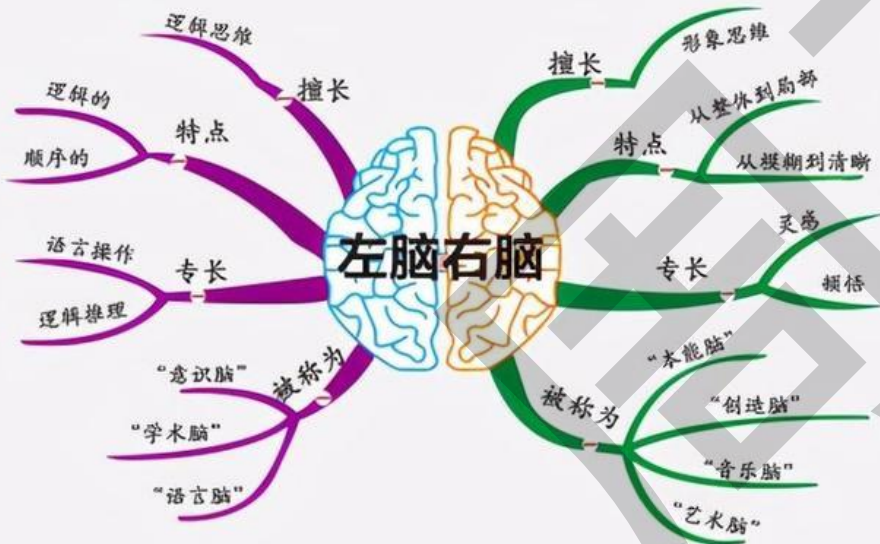


情感分析任务简介

认知智能

智商：理解与推理能力

情商：情感与表达能力

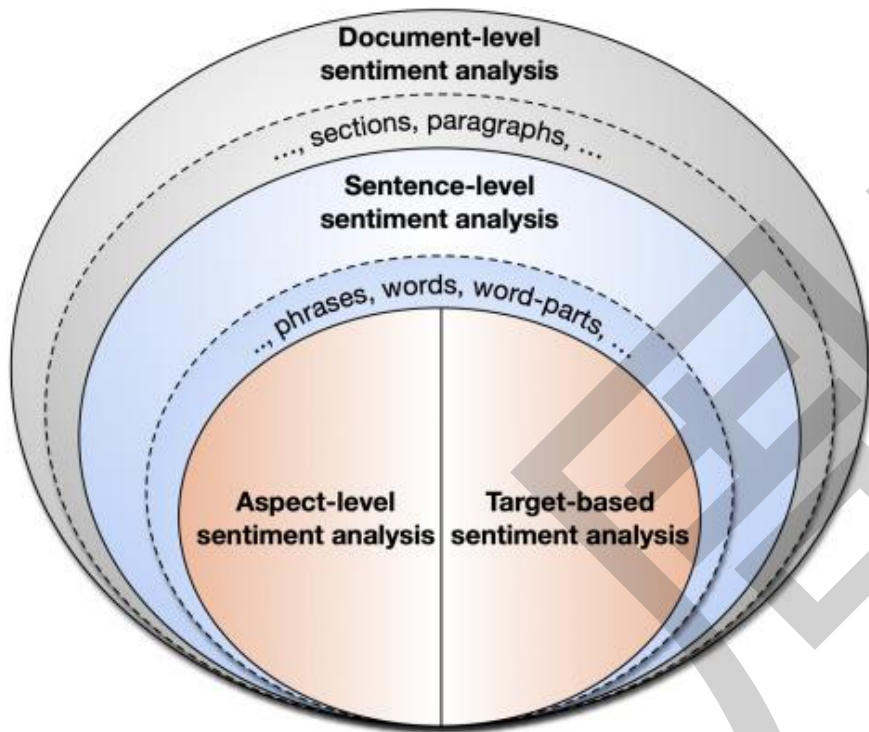


MIT Minsky (人工智能之父) 于1985年在“*The Society of Mind*”中指出，“问题不在于智能机器能否拥有任何情感，而在于机器实现智能时**怎么能够没有情感**。”



成功=80%情商+20%智商

情感分析任务简介



Sentiment Classification

The **ambience** was **nice**, but the **service** was **not so great**.

Aspect Term Extraction (AE):

- **ambience**
- **service**

Opinion Term Extraction (OE):

- **nice**
- **not so great**

Aspect-level Sentiment Classification (SC):

- **ambience** => **positive**
- **service** => **negative**

Aspect Term Extraction and Sentiment Classification (AESC):

- (**ambience**, **positive**)
- (**service**, **negative**)

Aspect-Oriented Opinion Term Extraction (AOE):

- **ambience** => **nice**
- **service** => **not so great**

Pair Extraction (Pair):

- (**ambience**, **nice**)
- (**service**, **not so great**)

Triple Extraction (Triple):

- (**ambience**, **nice**, **positive**)
- (**service**, **not so great**, **negative**)

Aspect-based Sentiment Analysis

情感分析任务简介

Document

Yesterday morning, a policeman visited the old man with the lost money, and told him that the thief was caught. The old man was very happy, and deposited the money in the bank.

Emotion Cause Extraction (ECE)

happy → a policeman visited the old man with the lost money

happy → and told him that the thief was caught

Emotion-Cause Pair Extraction (ECPE)

(The old man was very happy, a policeman visited the old man with the lost money)

(The old man was very happy, and told him that the thief was caught)

Trigger extraction

Argument extraction

Sentiment classification

Event	Polarity
[trigger 1] 发布 (release)	
[subject 1] 北玻股份 (Beibo co.)	
[object 1] 公告 (announcement)	Neutral
[time 1] 4月28日晚间 (On the evening of April 28)	
[loc 1] 投资者互动平台 (interactive platform for investors)	
[trigger 2] 增长 (increase)	
[subject 2] 营业收入 (operating income)	Negative
[object 2] 10%-20%	
[trigger 3] 增长 (increase)	
[subject 3] 净利润 (net profit)	Positive
[object 3] 91.28%	
[trigger 4] 同比下降 (year on year decrease)	
[subject 4] 净利润 (net profit)	Negative
[object 4] 64.65%	

Emotion Cause Extraction

Event-level
Sentiment
Analysis

Conversation
Emotion
Detection

Person A

U₁ I don't think I can do this anymore. [frustrated]

U₂ Well I guess you aren't trying hard enough. [neutral]

U₃ Its been three years. I have tried everything. [frustrated]

U₄ Maybe you're not smart enough. [neutral]

U₅ Just go out and keep trying. [neutral]

U₆ I am smart enough. I am really good at what I do. I just don't know how to make someone else see that. [anger]

Person B



Xia, Rui, et al. Emotion-Cause Pair Extraction: A New Task to Emotion Analysis in Texts. ACL, 2019

Hazarika, Devamanyu, et al. "Icon: Interactive conversational memory network for multimodal emotion detection.", EMNLP, 2018

Enhancing Event-Level Sentiment Analysis with Structured Arguments, SIGIR, 2022



華東師範大學
EAST CHINA NORMAL
UNIVERSITY

目录 | CONTENT

- 情感计算简介
- 预训练情感计算
- 生成式情感计算
- 情感计算发展趋势



基于预训练的情感分析

□ 如何将PLM提高情感分析任务性能？



Rich semantic information

Finetuning on sentiment analysis tasks

Sentiment **Knowledge-Enhanced** PTM

Continual Pre-training with Sentiment-aware Loss

主要使用抽取和分类方式

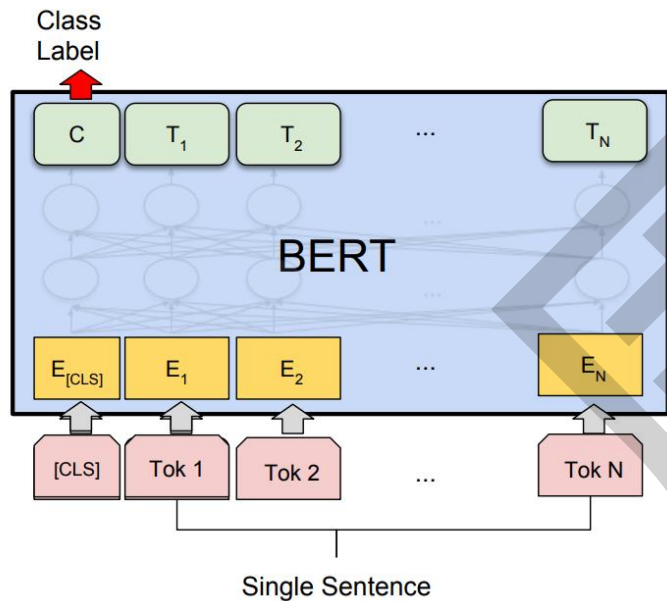
基于微调的情感分析

□ 直接当作一个分类模型进行微调 (情感分类)

利用PLM丰富的语义信息

和预训练的目标不一致

NLU的第一次打击



System	MNLI-(m/mm)	QQP	QNLI	SST-2
	392k	363k	108k	67k
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	90.4
OpenAI GPT	82.1/81.4	70.3	87.4	91.3
BERT _{BASE}	84.6/83.4	71.2	90.5	93.5
BERT _{LARGE}	86.7/85.9	72.1	92.7	94.9

基于微调的情感分析

- 将BERT用于ABSA任务 (ASC)
- 挑战: Aspect和句子的交互
 - 把ABSA转化为句子对分类
 - 与PLM的训练目标一致

Example:

LOCATION2 is central London so extremely expensive, LOCATION1 is often considered the coolest area of London.

QA-M:

What do you think of the safety of LOCATION1 ?

Methods	Output	Auxiliary Sentence	Model	Aspect			Sentiment	
				Acc.	F_1	AUC	Acc.	AUC
QA-M	S.P.	Question w/o S.P.	LR (Saeidi et al., 2016)	-	39.3	92.4	87.5	90.5
NLI-M	S.P.	Pseudo w/o S.P.	LSTM-Final (Saeidi et al., 2016)	-	68.9	89.8	82.0	85.4
QA-B	{yes,no}	Question w/ S.P.	LSTM-Loc (Saeidi et al., 2016)	-	69.3	89.7	81.9	83.9
NLI-B	{yes,no}	Pseudo w/ S.P.	LSTM+TA+SA (Ma et al., 2018)	66.4	76.7	-	86.8	-
			SenticLSTM (Ma et al., 2018)	67.4	78.2	-	89.3	-
			Dmu-Entnet (Liu et al., 2018)	73.5	78.5	94.4	91.0	94.8
			BERT-single	73.7	81.0	96.4	85.5	84.2
			BERT-pair-QA-M	79.4	86.4	97.0	93.6	96.4
			BERT-pair-NLI-M	78.3	87.0	97.5	92.1	96.5
			BERT-pair-QA-B	79.2	87.9	97.1	93.3	97.0
			BERT-pair-NLI-B	79.8	87.5	96.6	92.8	96.9

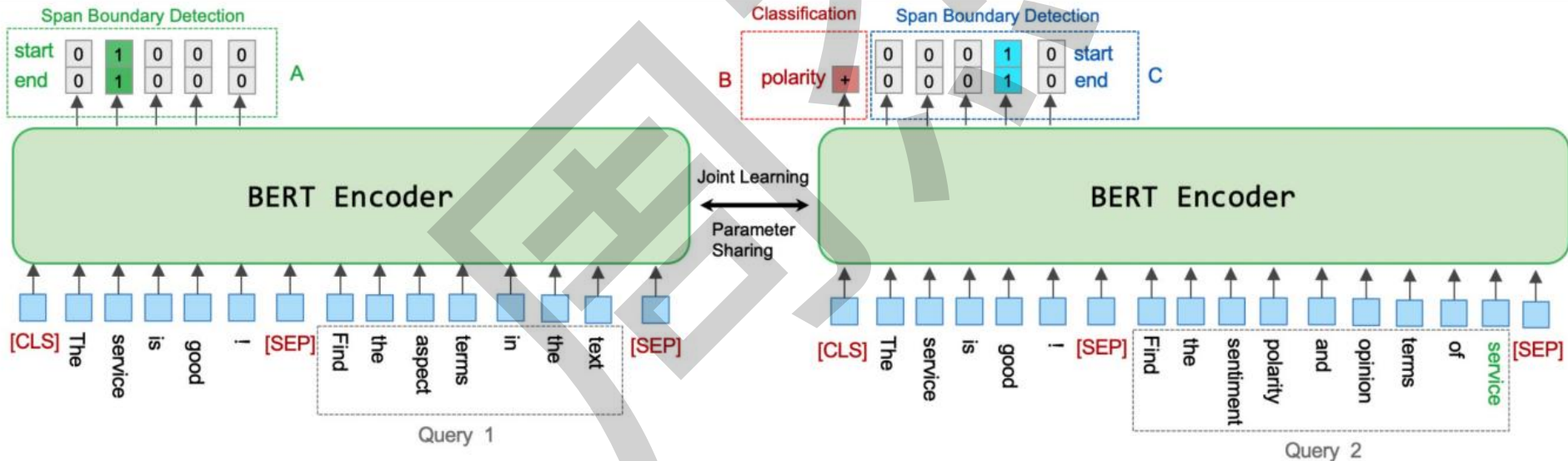
Table 2: The construction methods. Due to limited space, we use the following abbreviations: *S.P.* for *sentiment polarity*, *w/o* for *without*, and *w/* for *with*.

基于微调的情感分析

□ 将BERT用于ABSA任务 (Triple Extraction)

□ 挑战: Aspect和句子的交互

把ABSA任务转化为阅读理解任务



外部知识增强的情感分析

□ 结合外部情感知识用于情感分析

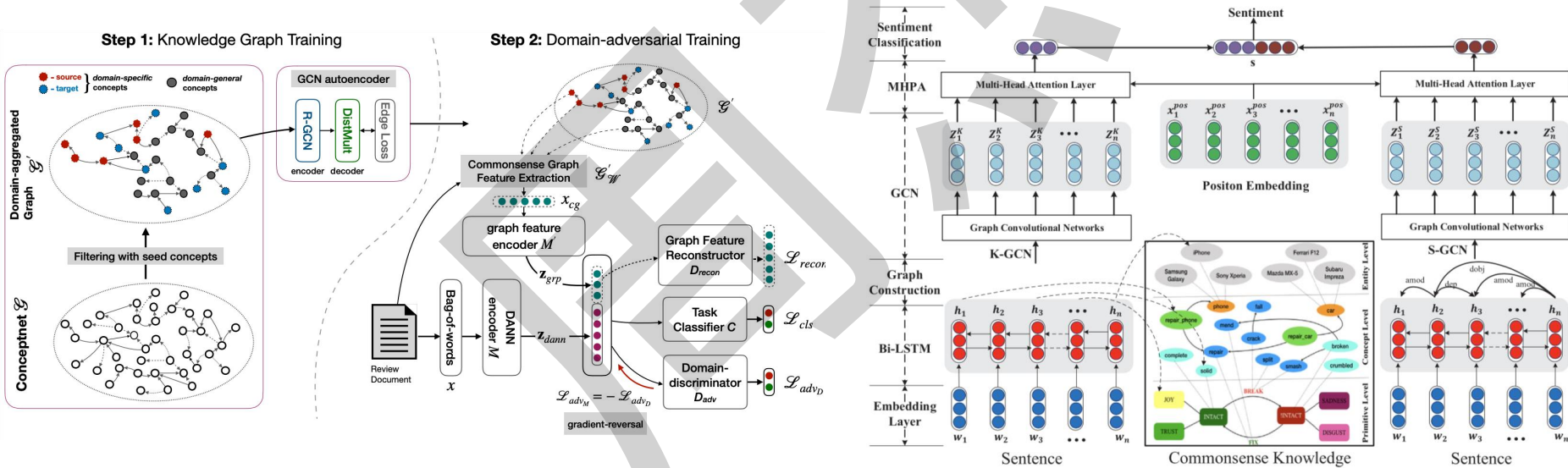
知识库，包括HowNet, ConceptNet、SenticNet

句法信息、词性信息

未来：GraphRAG

个人知识库

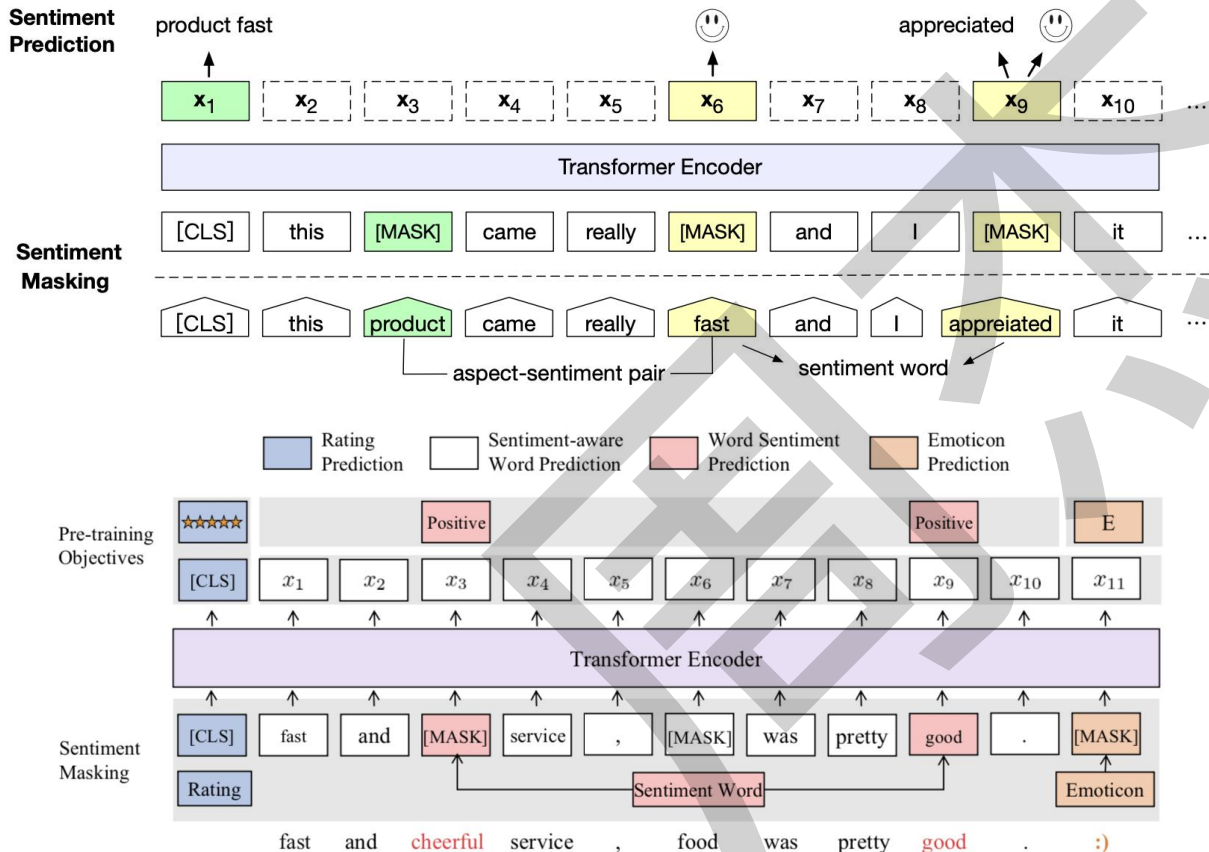
常识知识库（city 不 city 啊）



Ghosal, Deepanway, et al. "KinGDOM: Knowledge-Guided DOMain Adaptation for Sentiment Analysis." ACL, 2020.

Zhou, Jie, et al. "SK-GCN: Modeling Syntax and Knowledge via Graph Convolutional Network for aspect-level sentiment classification." KBS, 2020

基于预训练的情感分析



SKEP训练目标:

- 情感词预测
- 方面词预测

在14项中英情感分析典型任务上全面超越SOTA

SentiX训练目标:

- 情感词预测及其情感预测
- 字符表情预测
- 句子级别评分预测

仅仅10个样本就可以实现90+%的准确率

Tian, Hao, et al. "SKEP: Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis." ACL, 2020.

Zhou, Jie, et al. "Sentix: A sentiment-aware pre-trained model for cross-domain sentiment analysis." COLING, 2020.

知识增强的预训练情感分析

Linguistic Knowledge Acquisition from SentiWordNet

Context-aware Sentiment Attention

Term	POS	SN	Pscore / Nscore	Gloss (G)
good	a	1	0.75/0	Having desirable or positive qualities especially those suitable for a thing specified
good	a	2	0/0	Having the normally expected amount
...

$$\alpha_{good}^{(j)} = \text{softmax}\left(\frac{1}{SN_{good}^{(j)}} \cdot \text{sim}(\text{it is not a good movie}, G_{good}^{(j)})\right)$$

$$s(\text{good}, a) = \sum_j \alpha_{good}^{(j)} (\text{Pscore}_{good}^{(j)} - \text{Nscore}_{good}^{(j)}) = 0.62 > 0$$

$$\text{polar}(\text{good}, a) = \text{Positive}$$

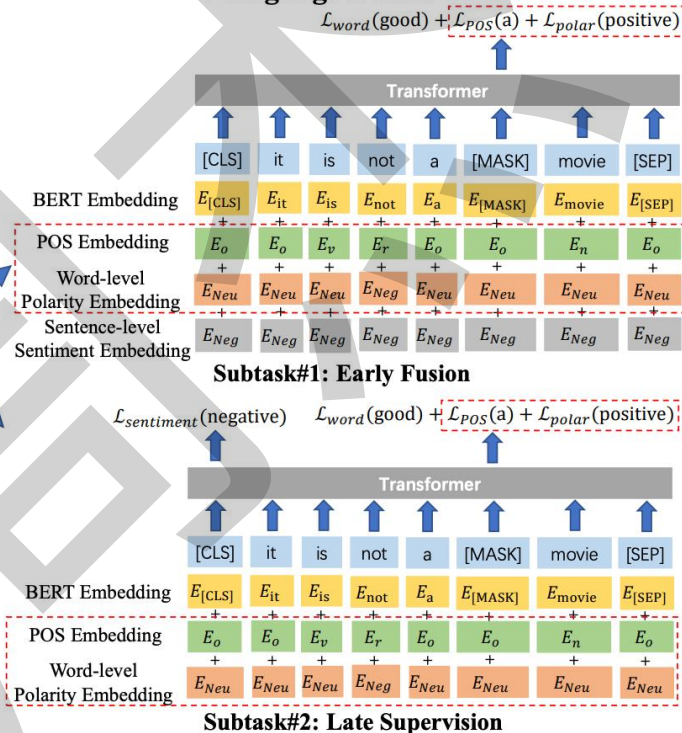
Part-of-Speech Tagging

Word	it	is	not	a	good	movie
POS Tag	o	v	r	o	a	n

Input: it is not a **good** movie

Pre-training Task: Label-aware Masked Language Model

$$\mathcal{L}_{word}(\text{good}) + \mathcal{L}_{pos}(\text{a}) + \mathcal{L}_{polar}(\text{positive})$$



SentiLARE

引入特征:

- POS
- 词级别情感
- 句子级别情感

训练目标:

- 情感词预测
- 情感词情感预测
- POS标签预测
- 句子情感预测



華東師範大學
EAST CHINA NORMAL
UNIVERSITY

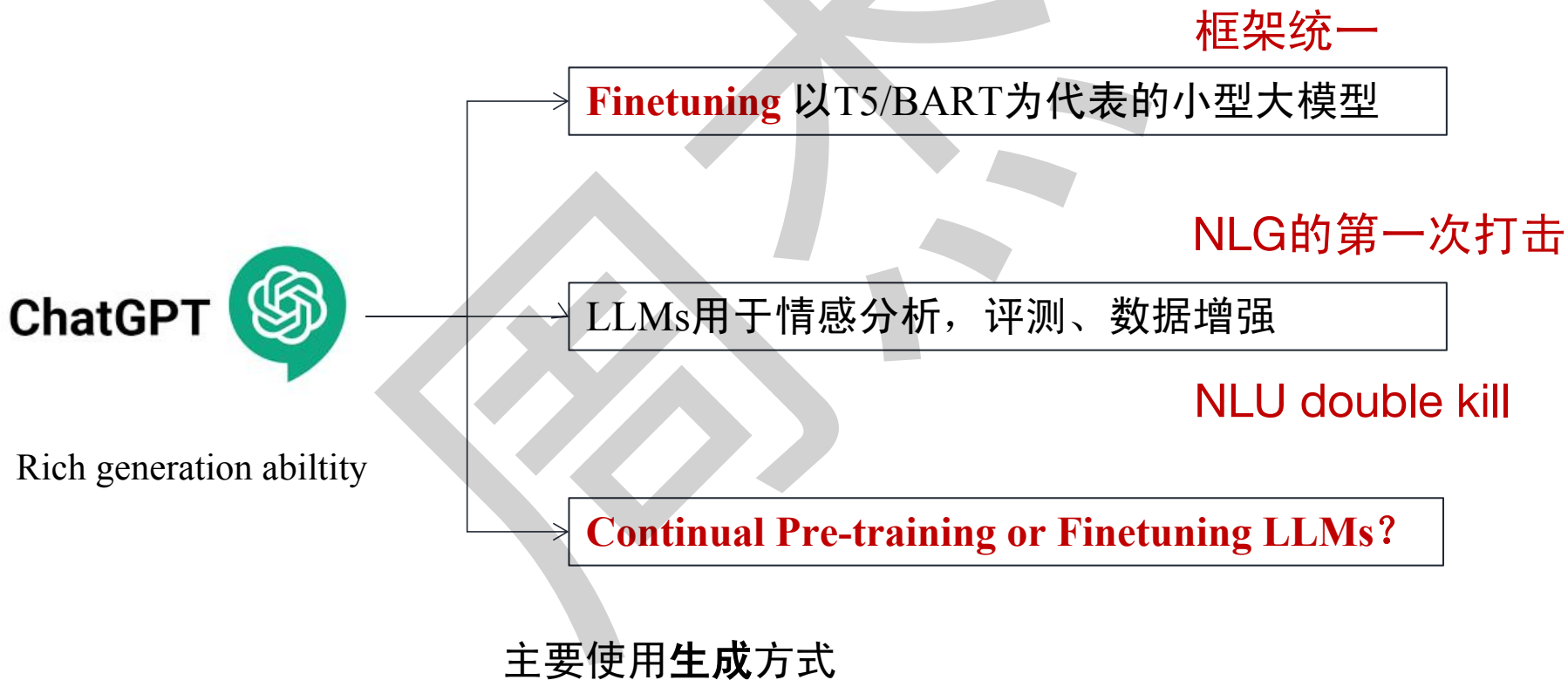
目录 | CONTENT

- 情感计算简介
- 预训练情感计算
- 生成式情感计算
- 情感计算发展趋势



生成式情感分析

□ 如何将GLM提高情感分析任务性能？



生成式情感分析

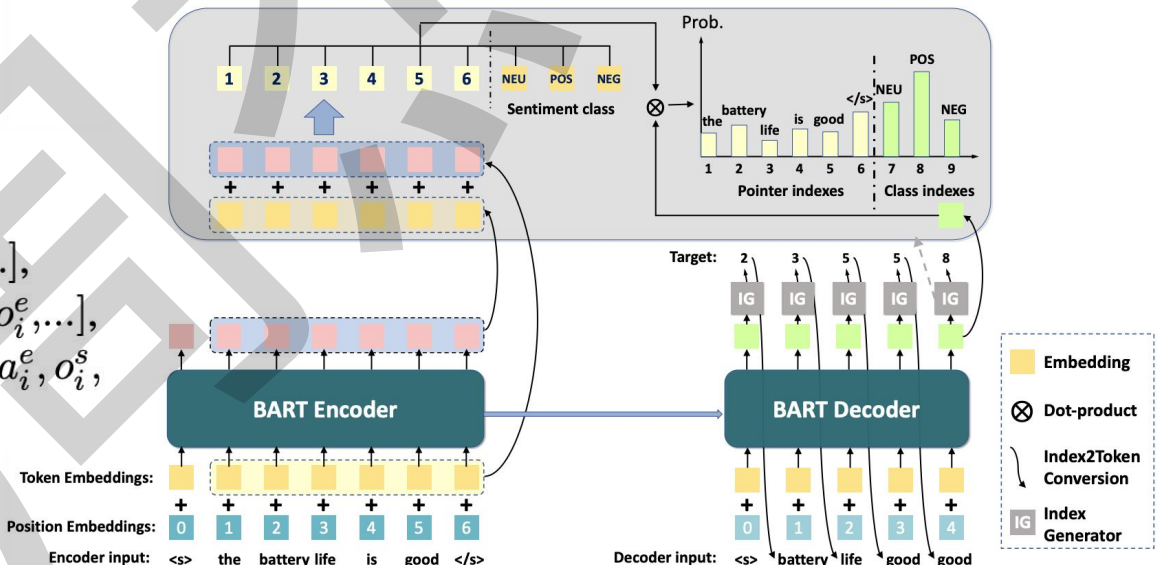
□ 基于指令学习的生成式情感分析

把所有ABSA子任务转化为生成任务

统一框架处理所有任务

使用**Pointer Network**

- $AE : Y = [a_1^s, a_1^e, \dots, a_i^s, a_i^e, \dots]$,
- $OE : Y = [o_1^s, o_1^e, \dots, o_i^s, o_i^e, \dots]$,
- $AESC : Y = [a_1^s, a_1^e, s_1^p, \dots, a_i^s, a_i^e, s_i^p, \dots]$,
- $Pair : Y = [a_1^s, a_1^e, o_1^s, o_1^e, \dots, a_i^s, a_i^e, o_i^s, o_i^e, \dots]$,
- $Triplet : Y = [a_1^s, a_1^e, o_1^s, o_1^e, s_1^p, \dots, a_i^s, a_i^e, o_i^s, o_i^e, s_i^p, \dots]$,



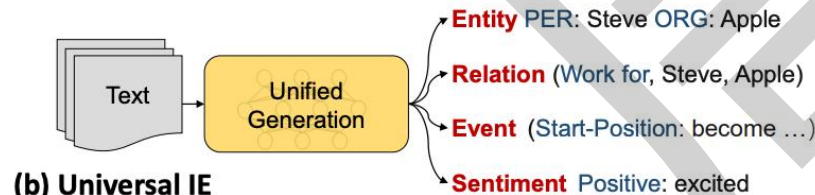
Yan, Hang, et al. "A Unified Generative Framework for Aspect-based Sentiment Analysis." ACL, 2021.

Scaria K, et al. Instructabsa: Instruction learning for aspect based sentiment analysis[J]. arXiv, 2023.

统一情感分析框架

Task	Schema	Instance								
Entity	PER: _ ORG: _	In 1997, Steve was excited to become the CEO of Apple.								
Relation	(_, Work for, _)	In 1997, Steve was excited to become the CEO of Apple.								
Event	<table border="1"> <thead> <tr> <th>Type</th> <th>Start Position</th> </tr> </thead> <tbody> <tr> <td>employee</td> <td></td> </tr> <tr> <td>employer</td> <td></td> </tr> <tr> <td>...</td> <td></td> </tr> </tbody> </table>	Type	Start Position	employee		employer		...		In 1997, Steve was excited to become the CEO of Apple.
Type	Start Position									
employee										
employer										
...										
Sentiment	Positive { Opinion: _; Target: _ }	In 1997, Steve was excited to become the CEO of Apple.								

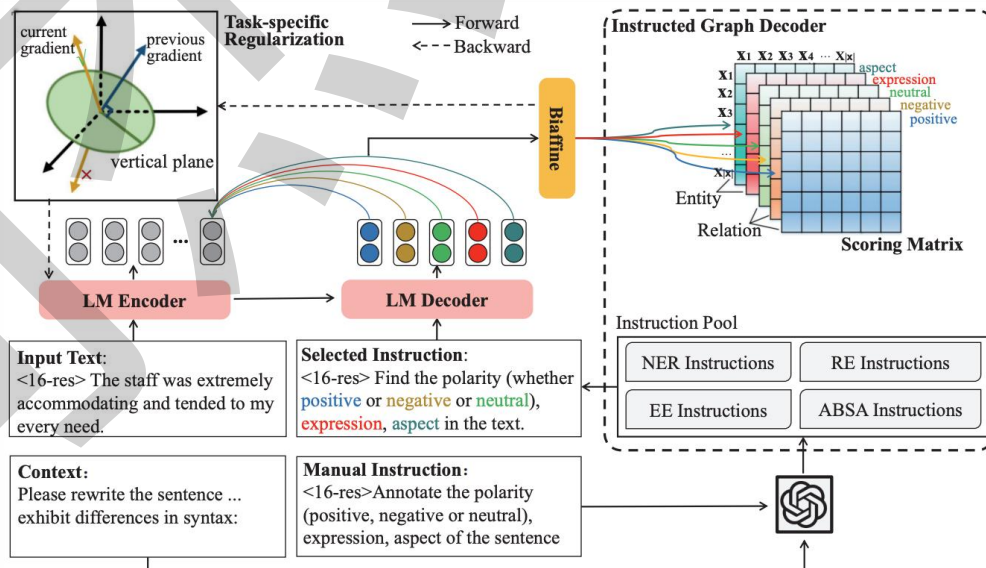
(a) Task-specialized IE



UIE 多任务训练

类似GPT到 InsturctGPT 演变

- 利用已有的结构化抽取任务数据集
- 模型泛化和迁移能力得到提升
- TIE解决数据标注不一致和冲突问题



TIE 多任务迁移

大模型情感分析能力评测

Task	Dataset	Baseline		LLM			SLM	
		random	majority	Flan-T5 (11B)	Flan-UL2 (20B)	text-003 (175B)	ChatGPT (NA)	T5 _{large} (770M)
<i>Sentiment Classification (SC)</i>								
Document-Level	IMDb	52.40	46.80	86.60	97.40	90.60	94.20	93.93
	Yelp-2	52.80	48.00	92.20	98.20	93.20	97.80	96.33
Sentence-Level	Yelp-5	19.80	18.60	34.60	51.60	48.60	52.40	65.60
	MR	47.40	49.60	66.00	92.20	86.80	89.20	90.00
Aspect-Level	SST2	49.20	48.60	72.00	96.40	92.80	93.60	93.20
	Twitter	34.20	45.40	43.60	47.40	59.40	69.40	67.73
	SST5	21.40	22.20	15.00	57.00	45.20	48.00	56.80
Average	Lap14	34.80	53.80	69.00	73.20	74.60	76.80	78.60
	Rest14	34.00	65.60	80.80	82.40	80.00	82.80	83.67
Average		38.44	44.29	62.20	77.31	74.58	78.24	80.65
<i>Aspect-Based Sentiment Analysis (ABSA)</i>								
UABSA	Rest14	NA	NA	0.00	0.00	47.56	54.46	75.31
	Rest15	NA	NA	0.00	0.00	35.63	40.03	65.46
	Rest16	NA	NA	0.00	0.00	40.85	75.80	73.23
	Laptop14	NA	NA	0.00	0.00	28.63	33.14	62.35
ASTE	Rest14	NA	NA	0.00	0.00	41.43	40.04	65.20
	Rest15	NA	NA	0.00	0.00	37.53	33.51	57.78
	Rest16	NA	NA	0.00	0.00	41.03	42.18	65.94
ASQP	Laptop14	NA	NA	0.00	0.00	27.05	27.30	53.69
	Rest15	NA	NA	0.00	0.00	13.73	10.46	41.08
	Rest15	NA	NA	0.00	0.00	18.18	14.02	50.58
Average		NA	NA	0.00	0.00	33.16	37.09	61.06
<i>Multifaceted Analysis of Subjective Text (MAST)</i>								
Implicit	Lap+Res	35.75	56.11	33.03	42.53	45.25	54.98	67.12
Hate	HatEval	48.00	36.31	56.09	70.80	67.79	50.92	46.94
Irony	Irony18	50.96	58.96	27.31	73.84	76.61	68.66	79.44
Offensive	OffensEval	46.67	41.86	32.78	74.44	73.31	64.88	80.76
Stance	Stance16	33.94	35.82	20.74	61.10	39.96	50.25	67.33
Comparative	CS19	49.36	73.89	54.46	85.67	74.52	75.80	89.49
Emotion	Emotion20	22.87	13.92	44.34	69.92	70.51	72.80	80.35
Average		41.08	45.27	38.39	68.33	63.99	62.61	73.05

Task	Datasets	#Test	Metric	Fine-tuned		Zero-shot	
				Baseline	SOTA	ChatGPT	+ Human
SC	SST-2	872	Acc	95.47 [†]	97.50 ^α	93.12	-
ABSC	14-Restaurant	1119	Acc / F1	83.94 [†] / 75.28 [†]	89.54 / 84.86 ^β	83.85 / 70.57	-
	14-Laptop	632	Acc / F1	77.85 [†] / 73.20 [†]	83.70 / 80.13 ^γ	76.42 / 66.79	-
E2E-ABSA	14-Restaurant	496	F1	77.75 [†]	78.68 ^δ	69.14	83.86
	14-Laptop	339	F1	66.05 [†]	70.32 ^δ	49.11	72.77
CSI	Camera	661	F1	93.04 [§]	-	74.89	-
CEE	Camera	341	F1	34.41 [‡]	-	9.10	51.28
ECE	Emotion Cause Dataset	100	F1	69.46 [‡]	-	74.01	-
ECPE	Emotion Cause Dataset	100	F1	65.20 [‡]	-	52.44	-

□ 简单的情感分类任务已经被大模型基本解决
ChatGPT can already serve as a universal and **well-behaved** sentiment analyzer.

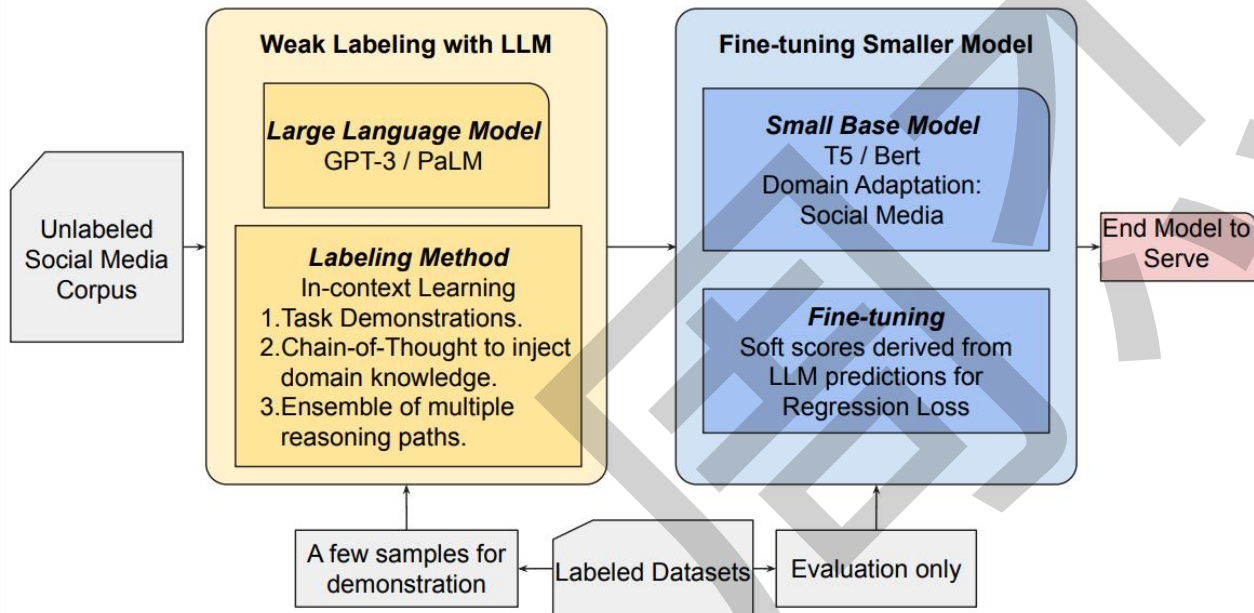
□ 细粒度的情感分析任务还有一定空间
比如ASTE, ECPE等任务

Zhang, Wenxuan, et al. "Sentiment Analysis in the Era of Large Language Models: A Reality Check." NAACL, 2024.

Wang, Zengzhi, "Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study", Arxiv, 2024

基于大模型数据增强的情感分析

用大模型来合成数据



大模型蒸馏

- 使用大模型生成弱标签数据
- 使用小参数模型进行微调

Only finetuning on T5\BART,
How about larger model ? ? ?



Pre-training Sentiment-aware LLMs??

Textual sentiment analysis is dead?



華東師範大學
EAST CHINA NORMAL
UNIVERSITY

目录 | CONTENT

- 情感计算简介
- 预训练情感计算
- 生成式情感计算
- 情感计算发展趋势



如何激发大模型的情感力

目前的情感机器人

- (1) 情感和情绪识别为主
- (2) 缺乏语音视觉的拟人度和表情，相对死板冷漠
- (3) 关注短时间的情绪

期待真的有情感的教育机器人

- (1) 具有同理心、情感丰富
- (2) 会说话、懂人，具有高情商
- (3) 长期陪伴，共同成长

◆ 多模态拟人能力有待提升

- (1) 声音、表情与语言的协调配合
- (2) 像人一样有“喜怒哀乐”的表现

◆ 懂心理和人工智能的结合

- (1) 懂心理、会说话
- (2) 关键时刻能够让人“豁然开朗”

◆ 拥有不一样的“有趣的灵魂”——自我

- (1) 不能千篇一律，每个机器应有自己的个性和独立性

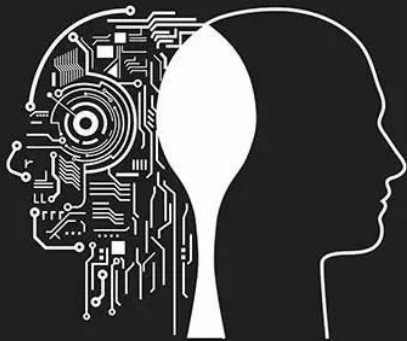


如何突破“画皮画虎难画骨，知人知面不知心”？



“类人”情感大模型

从文本情感感知到情感认知，情感交互的深度不断深入



2018年 GPT-1不能解决任何ToM任务；2022年GPT-3-davinci-002表现出与7岁儿童相当的水平；10个月后推出的 GPT-3.5-davinci-003则表现出与9岁儿童相当的水平

心智理论（Theory of Mind, ToM）是指个体理解他人具有不同心理状态的能力，这些心理状态包括信念、愿望、意图和情感。在人类社会互动中，心智理论至关重要，因为它允许我们**预测和解释**他人的行为，从而在复杂的社会环境中进行有效沟通和协作。

千古难题：女朋友说“我没有生气”，到底生气了没？



□ 深入结合心理学理论；

□ 构建更加复杂情感计算任务、数据集和评测指标

可持续学习的千人千面情感大模型

LLaMA V1.0



LLaMA V1.1



LLaMA V1.5



LLaMA V2.0



陪了三年的朋友还是三年前的那个朋友吗？

日久见人心

感情是可以培养的

白天对话

晚上进化

白天对话

晚上进化

白天对话

晚上进化

问题：现测得一个长方形四条边的长度分别为2、2、3、3厘米，求长方形的周长为___厘米。

这个问题主要涉及到长方形的性质，和周长的计算方法。你明白这两个概念吗？

对的，长方形的对角相等，然后周长是所有边的总和。

很好。现在就请你用这个概念来计算一下这个长方形的周长。

长方形的周长应该就 $2+2+3+3$ 的和，那么结果应该是9厘米。

真棒！ **专家纠正**
你的计算结果好像与实际不符，再仔细算一算看看是不是哪里出错了？

问题：现测得一个长方形四条边的长度分别为2、2、3、3厘米，求长方形的周长为___厘米。

这个问题主要涉及到长方形的性质，和周长的计算方法。你明白这两个概念吗？

对的，长方形的对角相等，然后周长是所有边的总和。

很好。现在就请你用这个概念来计算一下这个长方形的周长。

长方形的周长应该就 $2+2+3+3$ 的和，那么结果应该是9厘米。

真棒！ **专家纠正**
你的计算结果好像与实际不符，再仔细算一算看看是不是哪里出错了？

问题：现测得一个长方形四条边的长度分别为2、2、3、3厘米，求长方形的周长为___厘米。

这个问题主要涉及到长方形的性质，和周长的计算方法。你明白这两个概念吗？

对的，长方形的对角相等，然后周长是所有边的总和。

很好。现在就请你用这个概念来计算一下这个长方形的周长。

长方形的周长应该就 $2+2+3+3$ 的和，那么结果应该是9厘米。

真棒！ **专家纠正**
你的计算结果好像与实际不符，再仔细算一算看看是不是哪里出错了？

问题：现测得一个长方形四条边的长度分别为2、2、3、3厘米，求长方形的周长为___厘米。

这个问题主要涉及到长方形的性质，和周长的计算方法。你明白这两个概念吗？

对的，长方形的对角相等，然后周长是所有边的总和。

很好。现在就请你用这个概念来计算一下这个长方形的周长。

长方形的周长应该就 $2+2+3+3$ 的和，那么结果应该是9厘米。

真棒！ **专家纠正**
你的计算结果好像与实际不符，再仔细算一算看看是不是哪里出错了？

问题：现测得一个长方形四条边的长度分别为2、2、3、3厘米，求长方形的周长为___厘米。

这个问题主要涉及到长方形的性质，和周长的计算方法。你明白这两个概念吗？

对的，长方形的对角相等，然后周长是所有边的总和。

很好。现在就请你用这个概念来计算一下这个长方形的周长。

长方形的周长应该就 $2+2+3+3$ 的和，那么结果应该是9厘米。

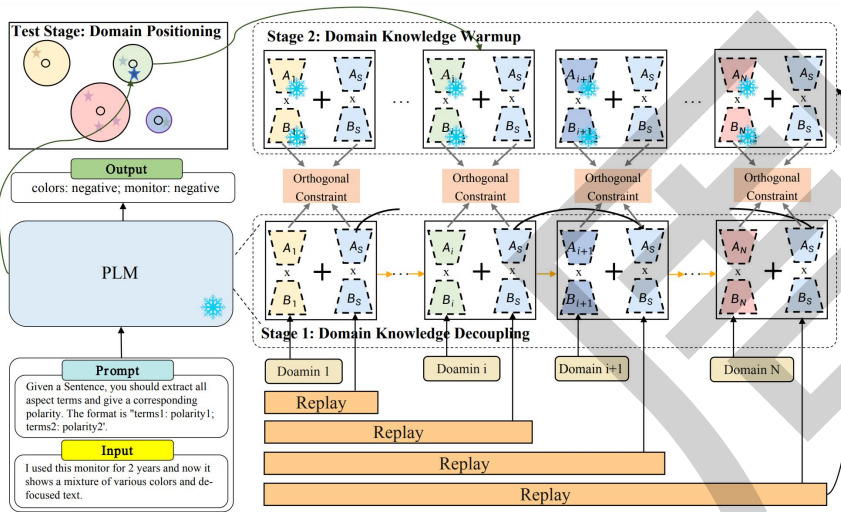
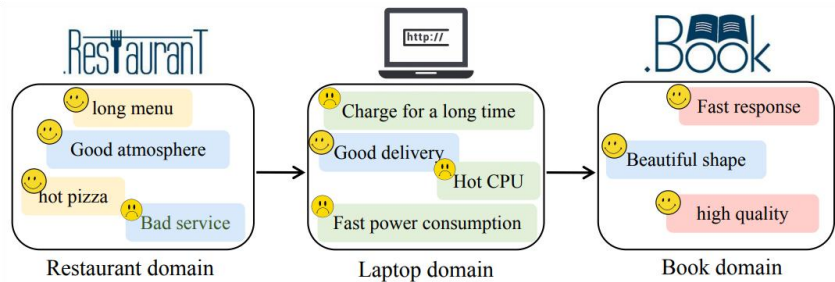
真棒！ **专家纠正**
你的计算结果好像与实际不符，再仔细算一算看看是不是哪里出错了？

大模型存在最大挑战：

- 缺少数据，尤其是过程交互数据
- 共用一个大模型，无法实现真正个性化

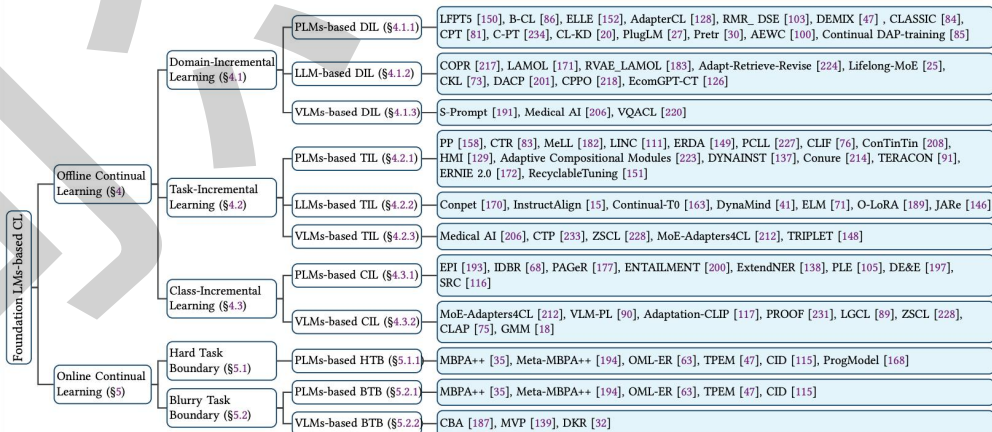
- 一边收集数据一边更新模型
- “长久陪伴”实现“一人一模型”
- 人机共同成长，实现“人机共生”

基于持续学习的情感分析



持续情感学习 (offline)

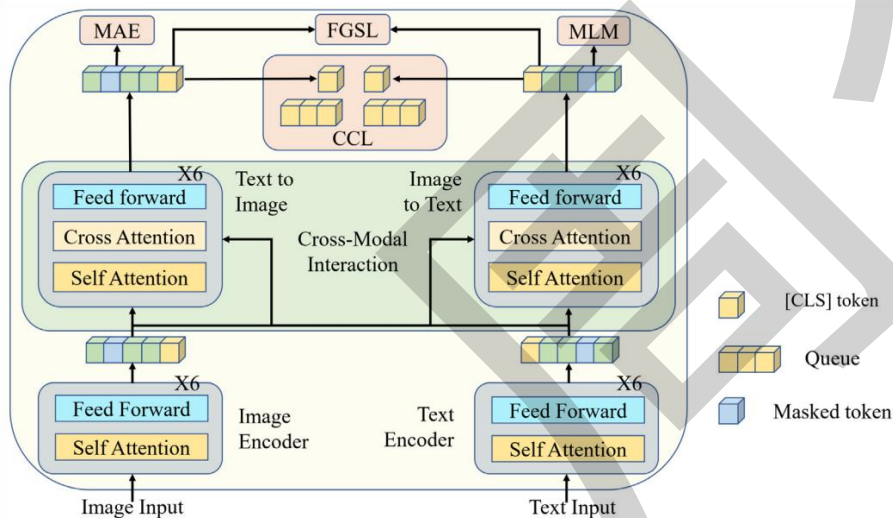
- Online持续学习快速精准更新
- 知识遗忘和知识迁移问题
- 判断是否旧类别，归纳新类别



大模型持续学习综述

多模态情感分析大模型

- 传统模型以输入文本、图像为主，更多模态（输出）
- 多模态情感数据集较少且质量较低
- 数字人+具身学习



Modality	Model	Twitter15		Twitter17		
		ACC	F1	ACC	F1	
Text	BERT	76.72±1.16	71.19±2.19	68.04±0.40	65.66±0.35	
	ResNet	57.65±1.00	32.52±2.66	57.79±0.99	51.98±1.23	
	ViT	59.65±1.13	31.25±2.71	59.53±0.95	54.08±0.78	
Image	Faster R-CNN	55.97±1.10	35.72±5.43	56.18±0.85	49.88±1.70	
	ResNet					
	Concatenate	75.29±0.45	68.71±1.34	67.92±0.56	65.32±0.53	
Tensor Fusion	74.19±0.94	68.93±0.57	66.66±1.21	63.99±1.61		
Self Attention	76.03±0.96	70.57±2.39	68.01±0.96	65.41±1.60		
Image2Text	77.13±1.33	71.48±1.90	69.37±0.36	66.85±0.79		
Text2Image	75.18±1.66	67.77±4.81	68.07±0.58	65.18±1.48		
Bi-direction	77.32±0.63	72.06±0.81	68.41±1.01	66.39±1.39		
ViT						
Multimodal	Concatenate	76.22±0.90	70.37±1.45	67.94±0.70	66.17±0.78	
	Tensor Fusion	73.44±0.78	67.46±1.45	65.46±1.67	62.02±1.40	
	Self Attention	75.08±0.41	68.94±0.83	67.52±0.58	65.56±0.35	
	Image2Text	77.11±0.44	71.91±0.42	69.14±0.52	66.96±0.68	
	Text2Image	75.12±1.01	69.40±1.38	67.52±1.06	64.49±1.46	
	Bi-direction	76.70±0.75	71.67±1.45	69.16±0.17	67.25±0.56	
	Faster R-CNN					
Concatenate	75.45±0.73	69.77±1.23	67.60±1.15	64.74±1.69		
Tensor Fusion	72.09±0.66	66.77±1.04	66.34±1.45	62.96±2.09		
Self Attention	76.09±0.89	70.08±1.37	68.09±1.10	66.12±1.23		
Image2Text	77.36±0.37	71.69±0.37	68.43±0.65	66.44±1.10		
Text2Image	70.82±2.99	57.94±5.81	60.31±6.43	54.50±7.06		
Bi-direction	76.57±0.46	70.88±0.89	69.51±0.62	67.50±0.37		

多模态情感分析大模型

- 传统模型以文本、图像为主，更多模态
- 多模态情感数据集较少且质量较低
- 数字人+具身学习

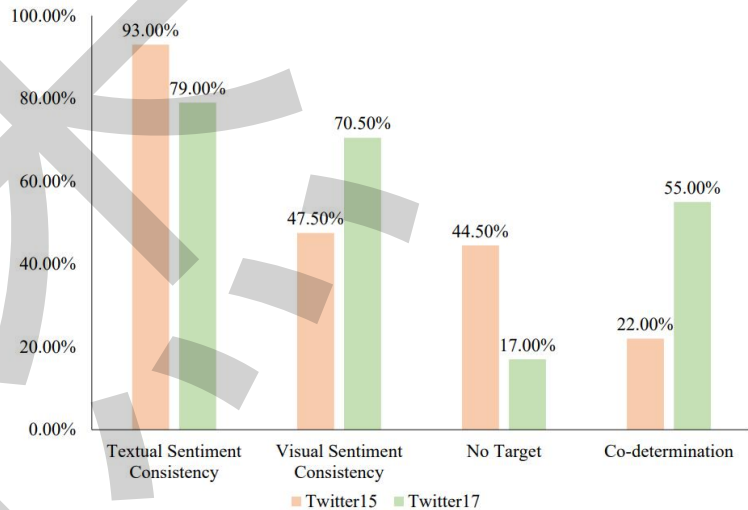
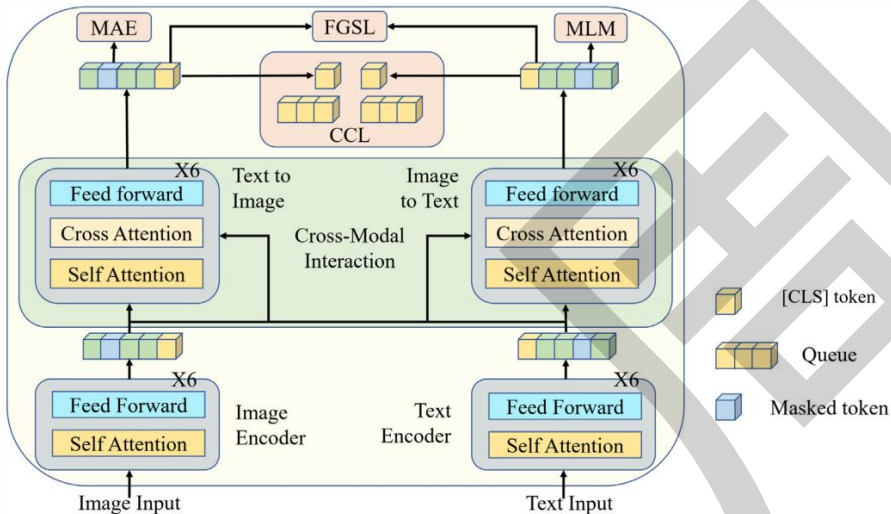
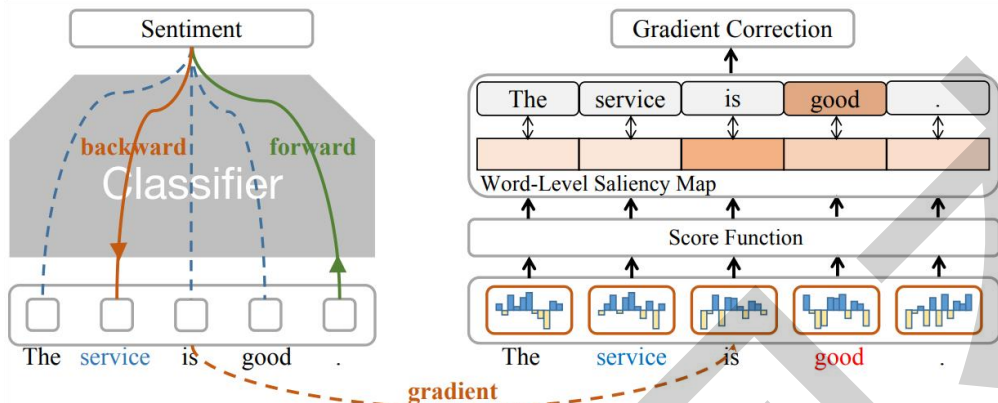


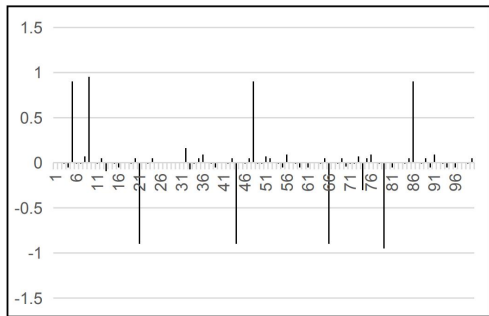
Figure 3: Annotation analysis. Textual/Visual Sentiment Consistency: the consistency of the target's sentiment in text/image with the sentiment in multimodal information. No Target: the percentage of images that are missing the target for sentiment analysis. Co-determination: the percentage of targets that sentiment is jointly determined by text and image.

基于大模型情感分析可解释性

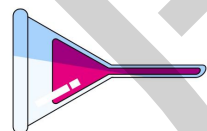


- 如何解释大模型情感背后的原因？
- 情感和哪些神经元有关？
- 如何控制情感？

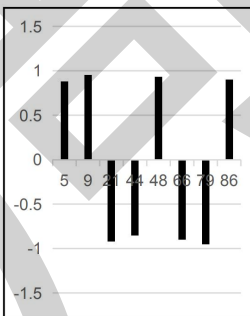
Gradient Distribution of Token



Information Bottleneck



IBG



谢谢!