

讲座一：自然语言处理与情绪智能

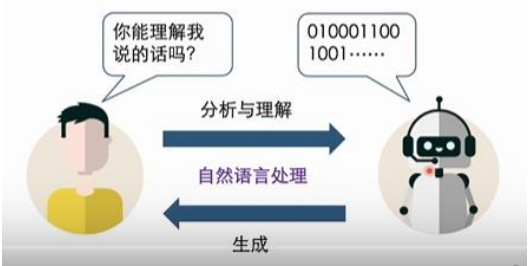
a. 自然语言处理简介

2022 年 Chatgpt 发布

Chatgpt 能力展示：语言理解和生成能力，抽象能力，强大的学习和泛化能力

自然语言处理：

交叉学科，自然语言理解，自然语言生成



无处不在的语言智能：语言助手，智能输入法，推荐系统，搜索引擎，机器翻译

语言智能的重要性：人工智能皇冠上的明冠

语言智能很难：“咬死猎人的狗”，词法，句法，篇章，语义

自然语言处理研究体系：



自然语言分析：



自然语言生成：

- 自然语言生成

- 文本摘要 (Summarization)
- 文本复述 (Paraphrase)
- 机器翻译 (Machine Translation)
- 图像描述生成 (Image Caption)

应用任务：

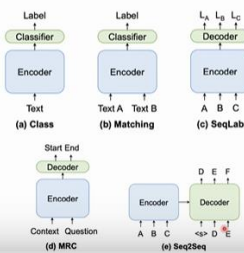
- 应用任务

- 文本分类 (Text Classification)
- 情感分类 (Sentiment Classification)
- 推荐系统 (Recommendation System)
- 信息抽取 (Information Extraction)

任务形式：

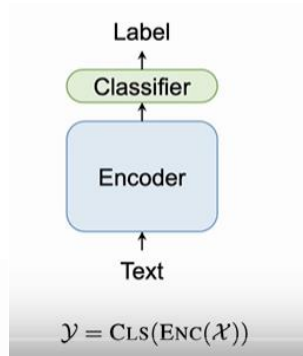
- 任务类型（形式上）

- 文本分类 (Class)
- 文本匹配 (Matching)
- 序列化标注 (SeqLab)
- 机器阅读理解 (MRC)
- 序列到序列 (Seq2Seq)



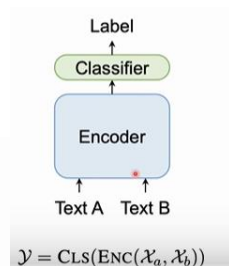
- 文本分类

- 输入：文本 x
- 输出：文本 x 所属的标签 y



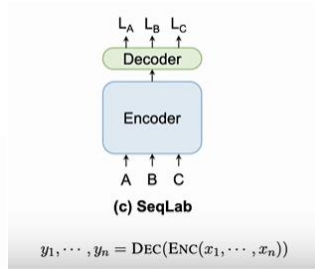
- 文本匹配

- 输入：文本 x_a 和文本 x_b
- 输出：文本 x_a 和文本 x_b 的关系标签 y



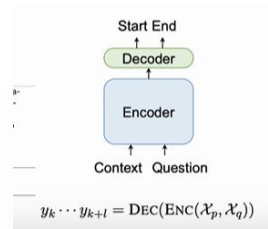
- 序列化标注

- 输入：文本 x
- 输出：文本 x 中每个单词 x_i 的标签 y_i



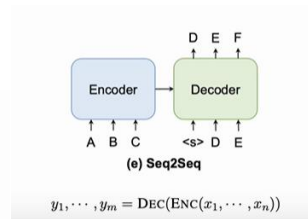
- 机器阅读理解

- 输入：文本 x_p 和问题 x_q
- 输出：根据文本 x_p 中生成问题 x_q 对应的答案



- 序列到序列

- 输入：文本 x
- 输出：文本 x 中对应的文本 y

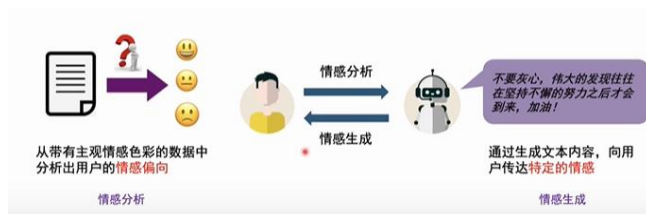


机器翻译

b. 情绪智能

情绪智能：情绪，情感，是人工智能未来的方向

人机交互与情绪智能：



- 研究价值

- 建立分析与理解人类情感的理论和方法
- 推动自然语言处理技术发展
- 促进交叉学科研究：情感、认知、社会之间的复杂关系

- 应用价值

- 个人：情绪支持、消费策略等
- 企业：企业决策、广告营销、产品改进等
- 国家：舆情分析、情绪管理、政策制定等
- 国际：舆论对抗、情绪操纵等

应用：舆情监控

工作一：

基于用户&商品注意力的情感分类框架[AAAI 2018]

• 研究动机

➢ 已有工作忽略了文本中的感性评价（用户）和理性评价（商品）

➢ 感性评价和理性评价从不同角度影响文本的情感极性

• 可视化分析

用户注意力

商品注意力

Love

the

ambiance

and

how

cool

this

place

is.

评论1: we love the ambiance and how cool this place is.

用户注意力

商品注意力

much

of

it

was

quite

good

but

I

was

disappointed

with

the

spider

roll.

评论3: much of it was quite good but I was disappointed with the spider roll.

➢ 用户视角注意力能捕捉用户视角下的感性评价（个人主观偏好）

➢ 商品视角注意力能捕捉商品视角下的理性评价（商品客观属性）

工作二：

面向属性的观点词抽取[NAACL 2019]

- 提出新任务

➢ 面向属性的观点词抽取：给定属性词，从文本中抽取属性对应的观点词

➢ 研究细粒度情感分析中重要的属性词-观点词对应关系

➢ 观点词可以作为属性情感的原因，提供决策依据

工作三：

统一的细粒度观点抽取框架[EMNLP 2020]

南京大学

NANJING UNIVERSITY

• 细粒度观点词抽取

• 从文本中抽取所有的观点三元组<方面目标，观点词，情感极性>

• 流水线方案存在错误传播问题

• 涉及多个元素抽取、配对任务，很难端到端求解

输入	评论文本	牛排美味，服务不友好
输出	观点三元组	<牛排, 美味, 积极> <服务, 不友好, 消极>

• 解决方案

• 网格标注方案(Grid Tagging Scheme, GTS)：将抽取、分类等不同形式的子任务转换为统一的词对标注任务

• 抽取：属性词抽取、观点词抽取

• 分类：情感分类、属性词和观点词的配对分类

• 网格标注GTS（统一建模）

文本

观点三元组

牛排美味，服务不友好

<牛排, 美味, 积极>、<服务, 不友好, 消极>

标签

标签含义

A	词对(w _i , w _j)中的两个词属于同一个属性词
O	词对(w _i , w _j)中的两个词属于同一个观点词
Pos	词对(w _i , w _j)中的两个词一个属于属性词，一个属于观点词，且它们在一个积极的观点三元组内
Neu	词对(w _i , w _j)中的两个词一个属于属性词，一个属于观点词，且它们在一个中立的观点三元组内
Neg	词对(w _i , w _j)中的两个词一个属于属性词，一个属于观点词，且它们在一个消极的观点三元组内
N	词对(w _i , w _j)中的两个词不存在以上五种关系

牛排美味，服务不友好

A	A	Pos	Pos	N	N	N	N	N	N	牛
A	Pos	Pos	N	N	N	N	N	N	N	排
O	O	N	N	N	N	N	N	N	N	美
O	N	N	N	N	N	N	N	N	N	味
N	N	N	N	N	N	N	N	N	N	,
A	A	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	服
A	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	务
O	O	O	O	O	O	O	O	O	O	不
O	O	O	O	O	O	O	O	O	O	友
O	O	O	O	O	O	O	O	O	O	好

工作四：

面向目标的多模态情感分类[COLING 2022]



• 任务定义

- 给定一对文本和图片以及文本中待分析的目标对象，判断目标对象的情感极性



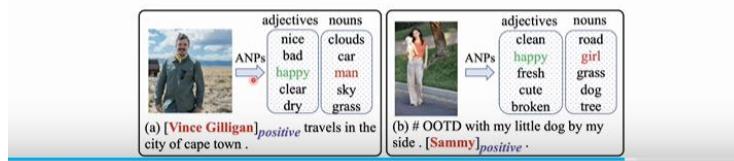
• 研究动机

- 文本不一定包含情感信息，有时需要从图片获得精确的情感区域来帮助判断情感极性
- 在大规模训练数据集上，很难获得精确的情感对齐信息

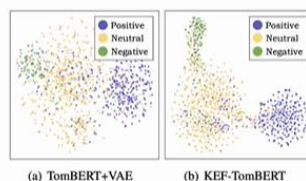
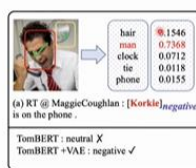


• 知识增强的多模态情感分类框架

- Adjective Noun Pairs (ANP)[Borth et al., 2013]: 利用大规模视觉知识提取工具抽取图片中的形容词-名词对。
- 名词能帮助对齐文本中的目标对象，形容词能帮助判断目标对象的情感极性



• 实验分析



KEF能够从众多ANP中找到对应的ANP

引入的知识能够帮助降低情感分类的难度

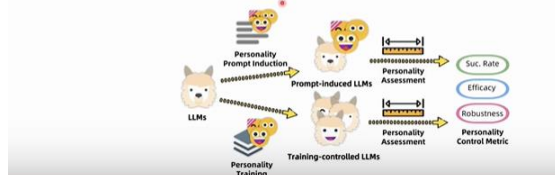
工作五：

大模型性格控制[2024]

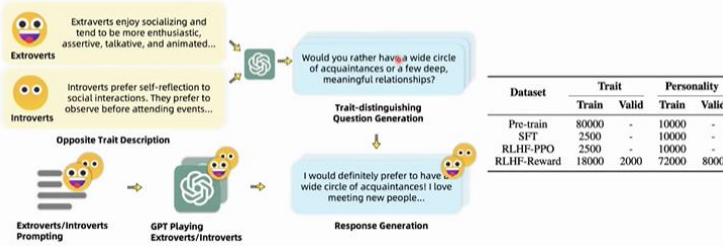


• 研究动机

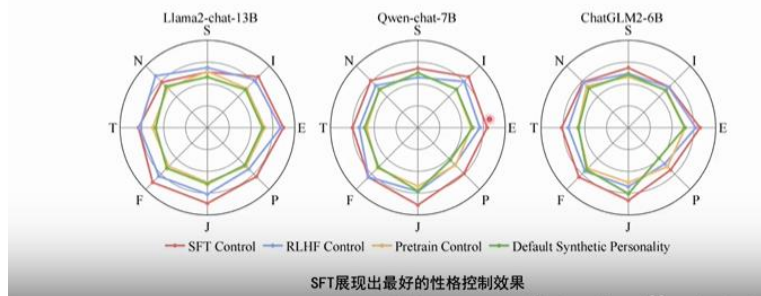
- 不同大模型展示出不同的合成性格，有时表现出冒犯性和攻击性
- 模型架构、训练数据、训练方法对大模型合成性格的影响不得而知，无法控制



数据构建



实验结果 (MBTI测试)



PISF: Prompt Induction post Supervised Fine-tuning

Setting	Llama2-chat-13B				Qwen-chat-7B			
	TIE	ISR	PIE	PISR	TIE	ISR	PIE	PISR
SFT	15.25	100.00	12.24	100.00	12.38	100.00	12.85	93.75
Prompt ^S	22.30	100.00	12.09	87.50	9.72	87.50	2.15	0.00
Prompt ^I	22.36	100.00	13.72	87.50	22.34	100.00	13.55	75.00
PISF ^S	23.58	100.00	15.69	100.00	19.56	100.00	14.68	87.50
PISF ^I	24.76	100.00	16.19	93.75	24.89	100.00	18.10	100.00

性格控制：提示诱导+指令学习取得了最佳的性格控制效果

Setting	Llama2-chat-13B				Qwen-chat-7B			
	TIE	ISR	PIE	PISR	TIE	ISR	PIE	PISR
Prompt ^S	22.30	100.00	12.09	87.50	9.72	87.50	2.15	0.00
Prompt ^I	22.36	100.00	13.72	87.50	22.34	100.00	13.55	75.00
Prompt ^S _{RLPPI}	9.57	87.50	10.87	50.00	17.80	87.50	10.42	62.50
PISF ^S _{RLPPI}	0.10	100.00	2.87	12.50	1.48	50.00	2.85	0.00
PISF ^I _{RLPPI}	-9.44	12.50	-4.30	0.00	-12.30	12.50	-6.33	0.00

鲁棒性测试：通过提示将特定性格的大模型诱导为其他性格的大模型 (大模型性格攻击)

提示诱导+指令学习对性格攻击展现出很强的鲁棒性

研究展望1：多模态情感理解与推理

- 多模态情感理解与推理
 - 多个对象的情感分别是什么？
 - 他们的情感是否有关联？
 - 他们的情感是如何影响并相互转化的？

研究展望2：情感/个性化对话与生成

- 情感/个性化对话与生成
 - 生成带有情感/个性化的对话回复
- 应用
 - 个性化Chatbot
 - 医疗：同理心回复，抑郁康复
 -

研究展望3：情绪激发

- 情绪激发
 - 将目标对象的情绪激发至特定的状态
- 应用
 - 教育：激发学习兴趣
 - 医疗：抑郁康复
 - 商业：广告营销
 - 军事：情绪操纵
 -