



南京大學
NANJING UNIVERSITY

RINC



句子级与片段级潜台词分析研究

- 答辩人：严 骅 MG1937028
- 导 师：申富饶 教授



目录

- 潜台词分析研究背景
- 研究内容
 - 中文潜台词数据集的构建与TAE score
 - 句子级分类任务：潜台词识别
 - 片段级分类任务：潜台词片段识别
- 实际应用：数据搜集与潜台词分析系统
- 研究生期间工作成果
- 总结与展望



研究背景

Research Background

研究背景



字典定义：在某一话语的背后，所隐藏着的那些没有直接、明白表达出来的意思。
简单而言，同一句话具有**字面意思**(不需要上下文就可以得到)和**隐含意思**(需要上下文才能得到)，且**字面意思和隐含意思的指向不一致**，即可以判定成具有潜台词。用数学描述为 $f(x) \neq f(x|c)$ 。

你的心有一道墙 老子翻过这道墙



我的心有一片海，你怎么办？



字面意思：心里没有墙，有海

隐含意思：**你放弃吧**

笨蛋才恋爱，我好想当笨蛋啊。



大笨蛋前来报道！



字面意思：我是大笨蛋

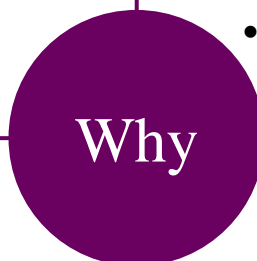
隐含意思：**我谈恋爱了**



1. 可以让**机器翻译更准确**

不同语种之间的表达习惯可能是不一样的，他们的交谈方式也通常都包含了潜台词。读懂了潜台词，可以让机器翻译更加准确。

- I burn you in this debate.
- 我在这场辩论中烧了你。[直译]
- 我在这场辩论中让你难堪。[平行语料]
- 我在这场辩论中赢了你。[伴有潜台词]



2. 可以让**语义理解更加准确**

表达方式的委婉可能导致对语义的理解不准确。而可以准确识别潜台词能让意图理解的更加准确。

- 你唱的真好听，别再唱了。
- 正向情感。[没有潜台词]
- 负向情感。[具有潜台词]

- 4. 舆情分析
- 5. 机器阅读理解
- ...

3. 让**对话系统更加流畅**

不能理解话语中的潜台词，让对话常常戛然而止。如果可以正确理解潜台词，则可以让对话更加流畅。



第三章中文潜台词数据集构建

1. 数据搜集
2. 标注方法与数据质量评估 ($Kappa$ \TAE\ F_1)
3. 数据分析

第四章：句子级分类问题： 潜台词检测

1. 潜台词判断是什么
2. 解决方案
3. 实验及其分析

进阶

第五章-任务一：片段级分类 问题：潜台词片段识别

1. 什么是潜台词内容识别
2. 解决方案
3. 实验及分析

第六章数据搜集与潜 台词分析系统

潜台词检测：

1. 输入：<目标语句><sep><上下文信息>
2. 输出：有潜台词(1)或不确定(0)或无潜台词(-1)

潜台词片段识别：

1. 输入：<目标语句><sep><上下文信息>
2. 输出： (s_i, e_i, t_i)
 s_i :第i个片段的开头位置
 e_i :第i个片段的结尾位置
 t_i :第i个片的类型

任务异同点：

1. 输入相同
2. 输出不同
3. 粒度不同



研究内容(一)

Chinese Subtext Dataset & TAE Score 中文潜台词数据集与TAE Score

数据搜集



数据标注



质量评估



数据分析



数据搜集

- 微博、知乎、网易云、哔哩哔哩
- 共72,494条源数据



图1. 数据来源示例

数据标注

- 匿名化标注：去除用户ID和用户昵称。保留明星名字
- 两段式标注
- 去除重复数据和无效数据
- 共得8,843条数据

7种标签信息

表1 标注示例

No	subt	sarc	meta	exag	homo	emot	sent
1	1; 赚到手软。。。。;T+0 对用户更加不好	-1;-	-1;-;-	1; 赚到手软	-1;-;-	0	1
2	1; 大型内卷培养皿; 学校的竞争就很激烈	-1;-	1; 大型内卷培养皿; 培养大量类似人员的	-1;-	-1;-;-	2	0
3	1; 房价已经翻了好几番; 大部分人买不起房	-1;-;-	1; 现在收入... 还是大部分, 但是房价已经翻了好几番了 [惊喜]	-1;-	-1;-;-	2	0
4	-1;-;-	-1;-	1; 白嫖党; 伸手不花钱族	-1;-	-1;-;-	6	0

质量评估

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

p_0 : 其实就是ACC.

p_e : 其实就是对模型偏向性的惩罚。

	1	-1
1	A	B
-1	C	D

$$p_0 = \frac{A+D}{A+B+C+D}$$

$$p_e = \frac{(A+C)(A+B) + (D+C)(D+B)}{(A+B+C+D)^2}$$

1. 数据不平衡时，很大部分都不能被正确衡量可靠性
2. 最好在完全独立情况下使用

1. 不受数据分布影响
2. 适用于完全独立与非完全独立



TAE score: $T = \sum_i t_i$

$$t_i = \frac{\exp(agr_i - rad_i) - \frac{1}{e}}{e - \frac{1}{e}}$$

agr : 衡量数据样本的标注一致性

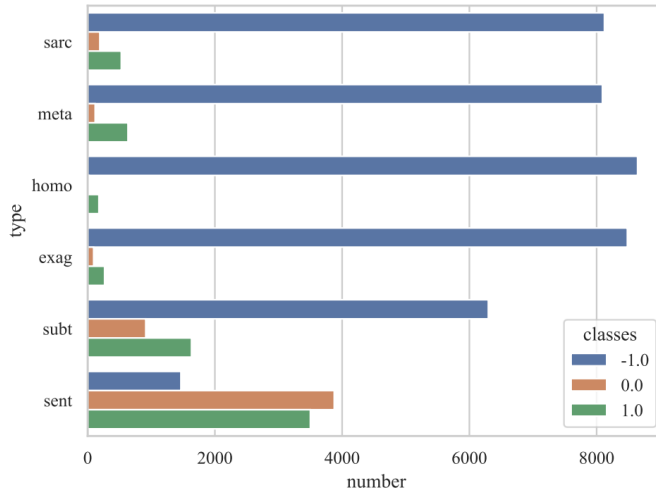
rad : 衡量数据样本的标注差异性

表2. 数据质量衡量结果

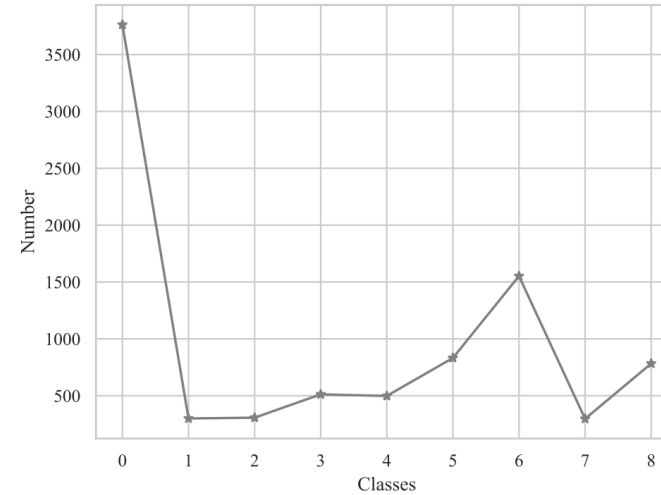
type	sarc	meta	subt	exag	homo	sent
Kappa	0.6028	0.6021	0.6042	0.7065	0.6113	0.7307
TAE	0.8071	0.5618	0.5016	0.8818	0.9515	0.5063
F_{1,SVM}	0.5087	0.4976	0.4657	0.5030	0.5322	0.5421

三个指标总体展示了我们所构建数据集是可靠可信的。

数据分析



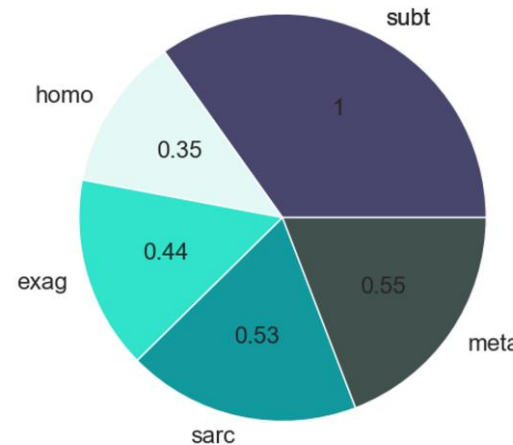
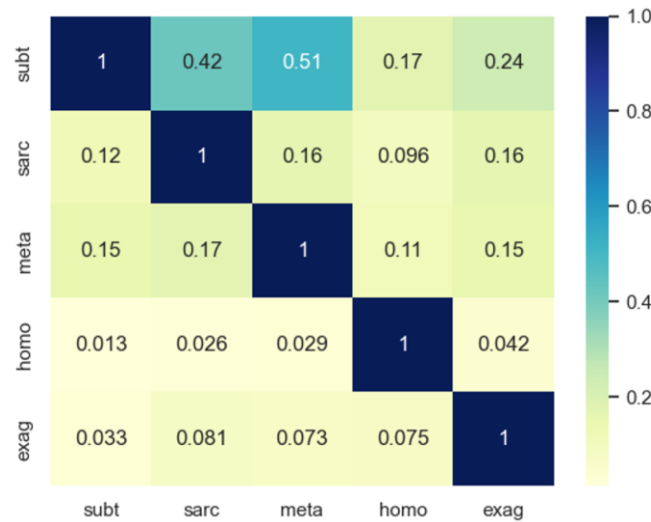
(a) 三类别标签的分布



(b) 情绪标签的分布

结论1: 数据分布是不平衡的。

图3. 数据标签分布



结论2: 不同类型的数据之间有 overlap, 而且不存在对称性

结论3: 不同的数据和subtext具有一定的关联性, 且metaphor和sarcasm与潜台词关联性较强。

图4. 数据的相关性分析



研究内容(二)

Sentence Level Classification

句子级分类任务：潜台词识别



输入: $S = \{ \langle cls \rangle, t_1, t_2, \dots, t_m, \langle sep \rangle, c_1, c_2, \dots, c_n \}$

输出: $y \in R$

Insight—以往模型的不足

1. 输入Self-Attention的embedding, 如果预先通过RNNs进行处理, 则 **embedding会变得相似**, 不利于提取Attention特征 (Tay et al. 2018.)
2. 使用Self-Attention, Token间的**Attention数值区分度不大**, 使得Attention的作用不如预期。

并行化抽取

Strengthen Attention

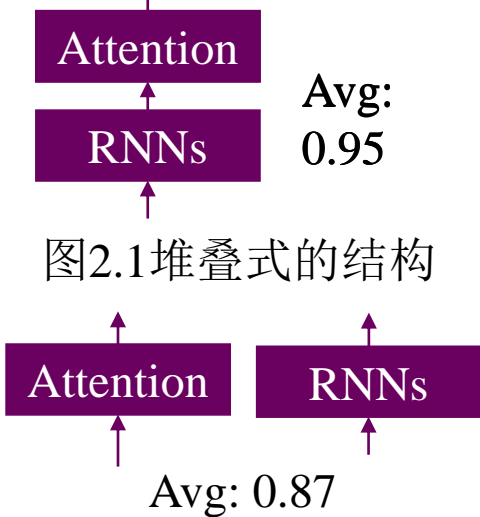


图2.1堆叠式的结构

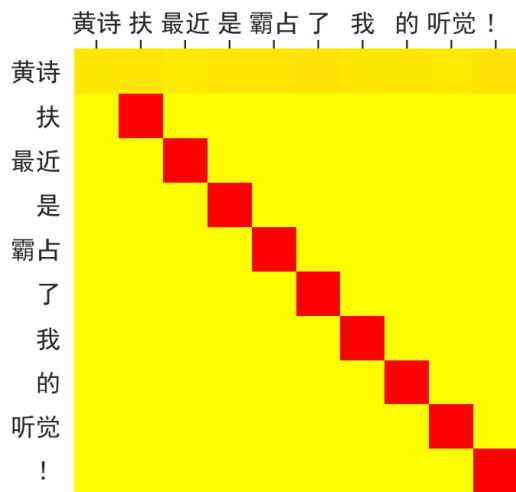


图2.3 Self-Attention注意力值

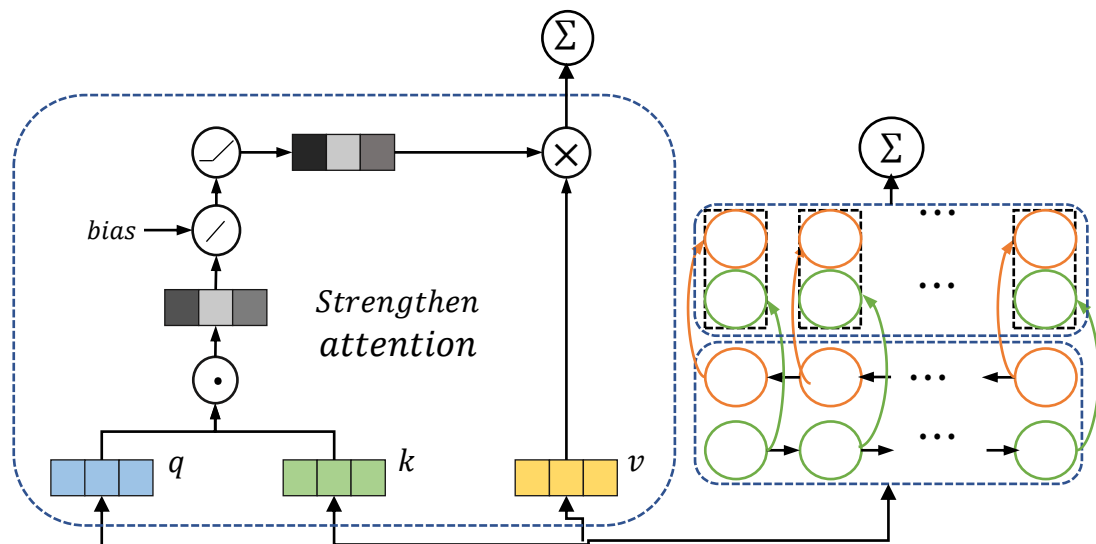


图2.2并行式的结构



输入: $S = \{ \langle cls \rangle, t_1, t_2, \dots, t_m, \langle sep \rangle, c_1, c_2, \dots, c_n \}$

输出: $y \in R$

Insight—潜台词检测的特点

- 3. Subtext 需要有足够的上下文和共识存在。本文研究主要建模是目标
语句和上文语句之间的语义变化关系 \longrightarrow 分别对comment和
comment + context建模
- 4. Subtext和Sarcasm、Metaphor具有更强的关联性。即意味着如果一个句子
具有比喻和反讽，则它有更大的可能性具有潜台词。 \longrightarrow 使用多任务
- 5. 数据分布是不平衡的：在loss上需要加以限制。 \longrightarrow 代价敏感分析

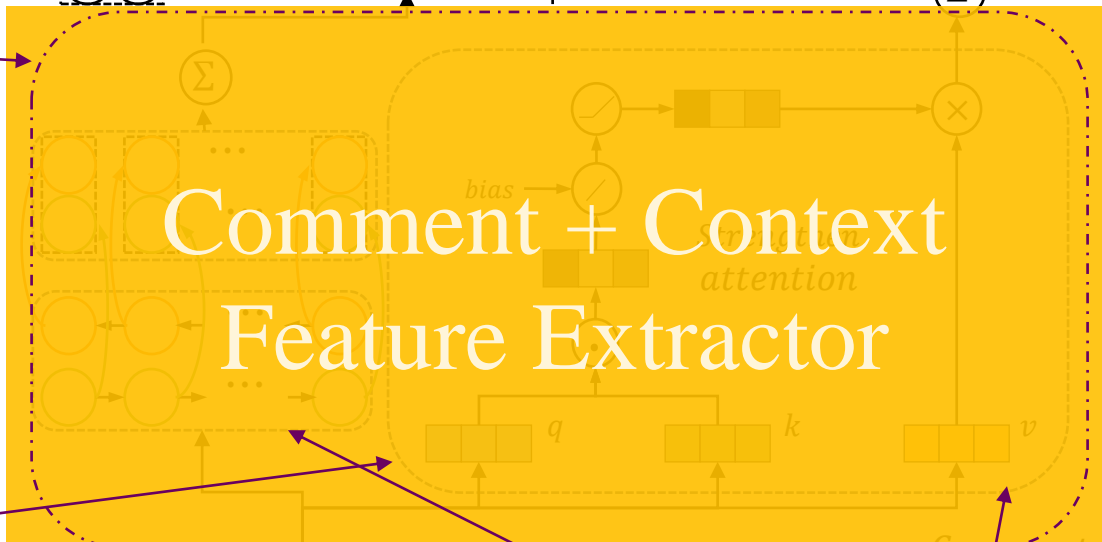
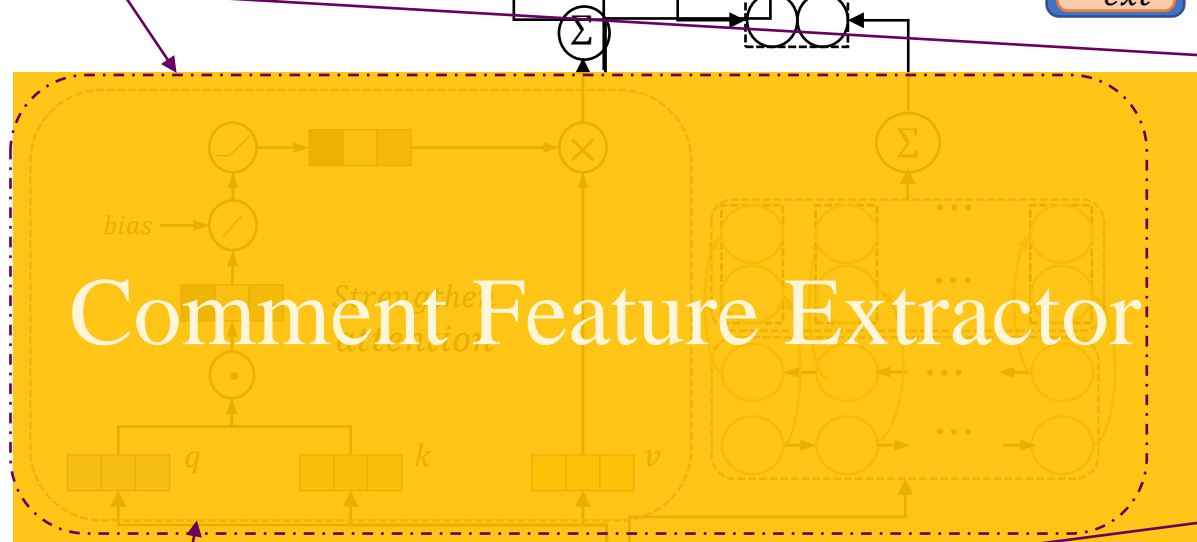
$$w_i = g\left(1 - \frac{\exp(f_i)}{\sum_{j \in \{-1, 0, 1\}} \exp(f_j)}\right) \longrightarrow \text{来源于验证集的结果}$$



SASICM模型结构

Insight 3. 双塔结构建模语义在上下文中的变换情况

Insight 4. 多任务建模相关关系



Insight 2. Strengthen Attention:使attention更具有区分度

Insight 1, 并行化抽取

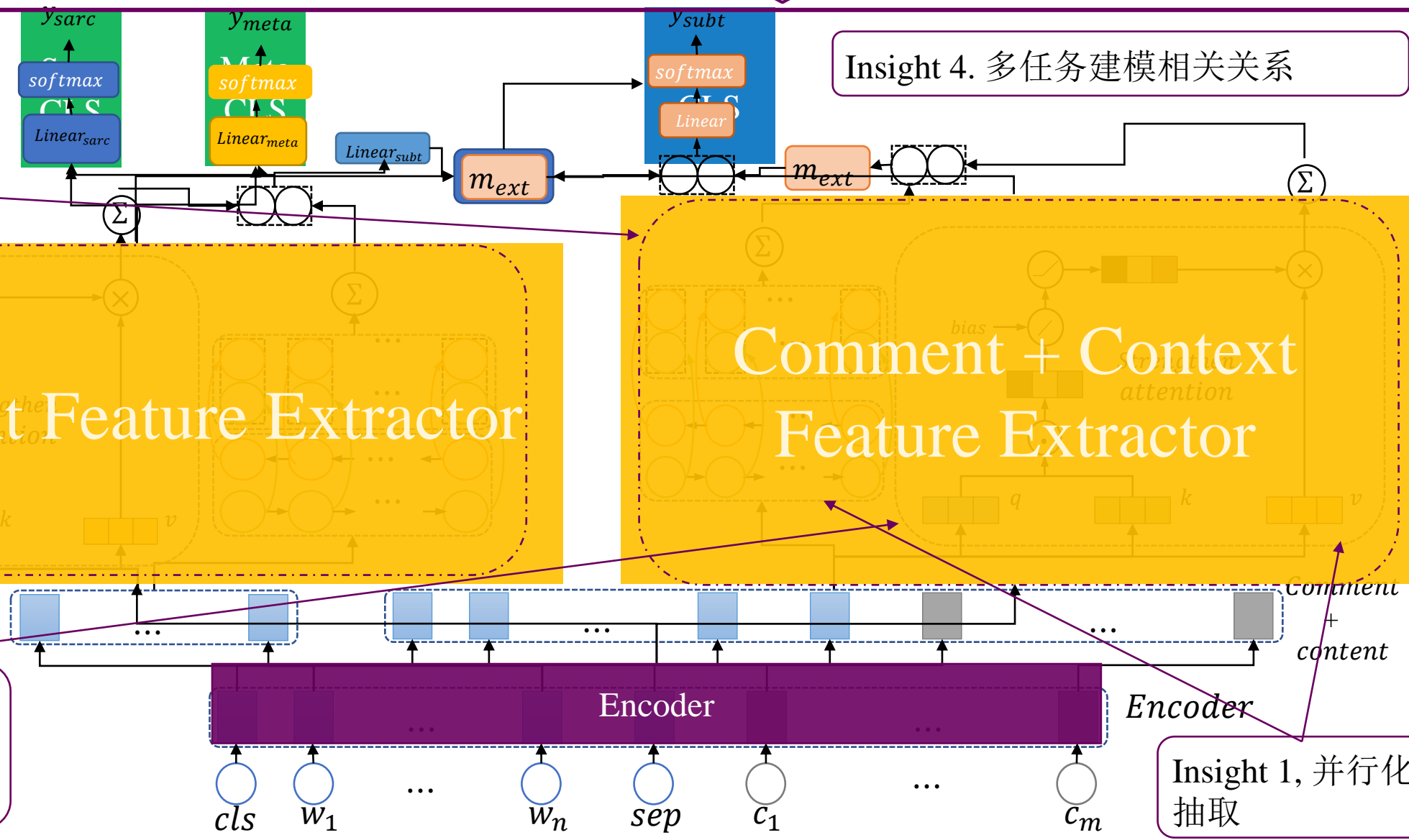


表 4-2: 三任务模型的实验结果。其中“p”为准确率,“r”召回率。下划线所表示的为我们的模型 (SASICM), $*_m$ and $*_s$ 分别代表反讽 (sarcasm) 和比喻 (metaphor) 的结果。

Model	Subtext Task				Metaphor Task				Sarcasm Task			
	p(%)	r(%)	F ₁ (%)	acc(%)	p _m (%)	r _m (%)	F _{1m} (%)	acc _m (%)	p _s (%)	r _s (%)	F _{1s} (%)	acc _s (%)
<u>SASICM_e</u>	67.57	68.52	67.04	68.52	83.72	91.50	87.43	91.50	84.38	91.86	87.96	91.86
<u>SASICM_{BERT}</u>	66.35	66.38	65.89	66.38	84.73	88.02	86.28	88.02	85.57	90.02	87.70	90.02
MIARN3	62.89	68.04	63.80	68.04	83.34	91.12	86.73	91.12	82.51	90.84	86.48	90.84
BTM3	63.24	67.33	63.08	67.33	84.25	90.26	86.89	90.26	82.62	91.08	86.43	91.08
BERT3	63.98	64.59	64.22	64.59	84.9	86.72	85.76	86.72	85.39	88.64	86.97	88.64
GBP	57.39	57.38	57.38	57.38	84.83	85.22	85.02	85.22	85.44	85.76	85.60	85.76
HP	81.05	76.82	78.20	76.82	92.20	89.54	88.65	89.54	93.01	92.90	92.89	92.89

表 4-3: 双任务框架的实验结果

Model	Subtext Task				Metaphor Task				Sarcasm Task			
	p(%)	r(%)	F ₁ (%)	acc(%)	p _m (%)	r _m (%)	F _{1m} (%)	acc _m (%)	p _s (%)	r _s (%)	F _{1s} (%)	acc _s (%)
<u>SASICMSS</u>	65.17	69.02	66.07	69.02	-	-	-	-	84.38	91.86	87.96	91.86
MIARNSS	62.15	71.61	63.94	71.61	-	-	-	-	85.33	91.14	87.87	91.14
BTMSS	61.66	71.31	62.52	71.31	-	-	-	-	84.98	92.19	87.82	92.19
BERTSS	61.97	72.09	64.41	72.09	-	-	-	-	84.98	92.19	87.82	92.19
<u>SASICMSM</u>	63.23	66.39	62.80	66.39	83.78	90.97	87.21	90.97	-	-	-	-
MIARNSM	61.75	70.65	63.68	70.65	84.62	91.27	87.62	91.27	-	-	-	-
BTMSM	61.57	71.71	62.27	71.71	84.35	91.72	87.85	91.72	-	-	-	-
BERTSM	61.97	72.02	64.40	72.02	84.32	91.82	87.91	91.82	-	-	-	-

表 4-4: 单任务框架的实验结果

Model	Subtext Task			
	p(%)	r(%)	F ₁ (%)	acc(%)
<u>SASICMSubt</u>	63.80	69.99	66.56	69.99
MIARNSubt	61.70	70.56	63.64	70.56
BTMSubt	62.04	71.96	62.59	71.96
BERT+FF	61.98	72.10	64.41	72.10
SVM	60.67	72.00	60.60	72.00
LR	55.93	72.05	60.44	72.05
MEC	51.98	72.10	60.40	72.10
NB	61.14	11.50	13.35	11.50
DT	62.19	66.62	63.09	66.62

1. 该问题可以解决, 且效果上而言传统方法不如深度学习方法
2. 三任务的效果最好, 说明更多的任务有助于缓解特征有偏的问题, 同时可以引入更多的线索。
3. SASICM的效果比其余模型具有极大的优势。



表 4-5: 消融实验结果

Model	Subtext Task				Metaphor Task				Sarcasm Task			
	p(%)	r(%)	F ₁ (%)	acc(%)	p _m (%)	r _m (%)	F _{1_m} (%)	acc _m (%)	p _s (%)	r _s (%)	F _{1_s} (%)	acc _s (%)
SASICMSt	67.68	69.13	66.61	69.13	83.72	91.50	87.43	91.50	84.38	91.85	87.96	91.85
SASICMG	65.12	68.93	66.23	68.93	83.72	91.50	87.43	91.50	84.38	91.86	87.96	91.86
SASICMWC	64.46	70.07	66.25	70.07	83.32	91.69	87.30	91.69	84.38	91.85	87.95	91.85
SASICMSL	66.07	68.53	66.28	68.53	83.72	91.50	87.43	91.50	84.38	91.85	87.96	91.85
SASICMSA	65.47	67.94	65.16	67.94	84.58	91.00	87.40	91.00	84.68	89.35	86.82	89.35

SASICMSt: 使用堆叠结构组合RNNs和Attention

SASICMG: 使用GRU替换LSTM

SASICMWC: 不对Strengthen Attention的参数进行限制

SASICMSL: 使用单项的LSTM特征

SASICMSA: 使用Self-Attention

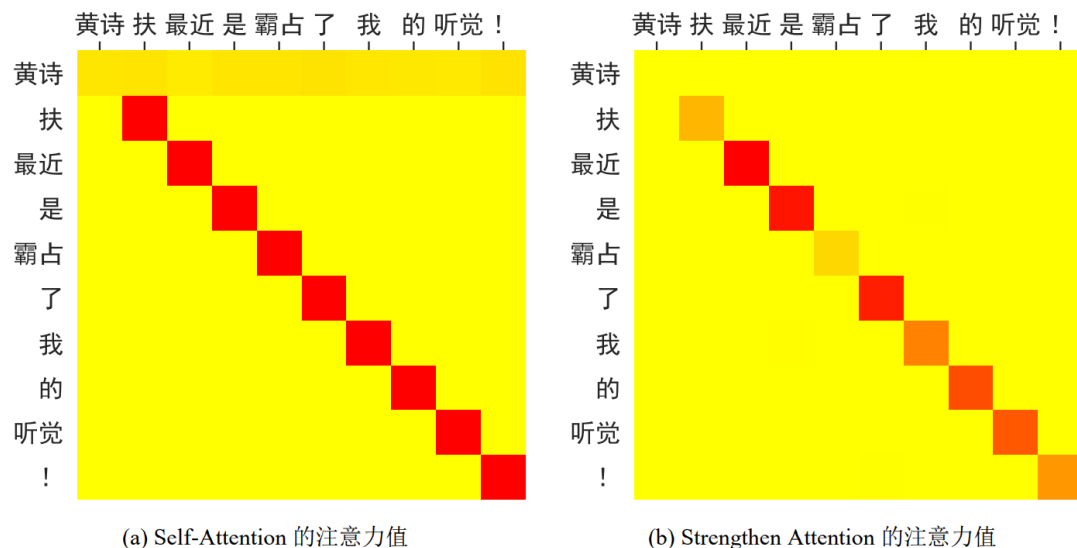


图 4-4: 自注意力机制 (Self-Attention) 和强化注意力机制 (Strengthen Attention) 的对比

结论:

1. Strengthen Attention比Self Attention学习到的Attention结果更有分辨度。
2. 对Strengthen Attention需要加以限制，否则，其学习到的缩放因子和阈值可能与初衷背离。
3. 并行式的结构更有利于提取句子的不同类型特征
4. 在潜台词识别任务上，LSTM要更好。



研究内容(三)

Segment Level Classification

片段级分类任务：潜台词片段识别



输入: $S = \{ \langle cls \rangle, t_1, t_2, \dots, t_m, \langle sep \rangle, c_1, c_2, \dots, c_n \}$

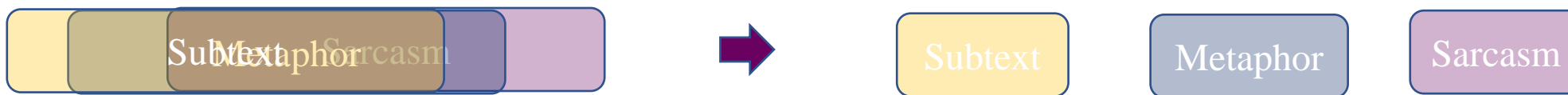
输出: $\{ (s_1, e_1, t_1), (s_2, e_2, t_2), \dots, (s_k, e_k, t_k) \}, s_i, e_i \in R, t_i \in \{ subtext, sarcasm, metaphor \}$



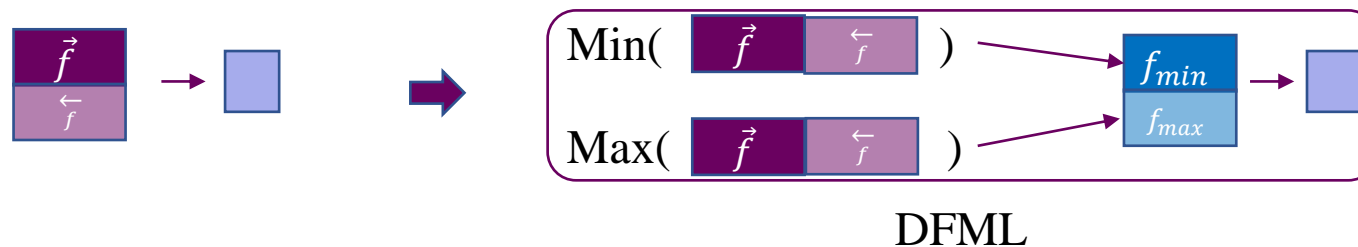
与命名实体识别具有同等的形式。

Insight——模型结构的设计

1. 基于普通序列标注的方法，难以解决嵌套问题。



2. 直接采用拼接的方法建模上下文特征，难以有效抽取上下文特征





输入: $S = \{ \langle cls \rangle, t_1, t_2, \dots, t_m, \langle sep \rangle, c_1, c_2, \dots, c_n \}$

输出: $\{ (s_1, e_1, t_1), (s_2, e_2, t_2), \dots, (s_k, e_k, t_k) \}, s_i, e_i \in R, t_i \in \{ \text{subtext}, \text{sarcasm}, \text{metaphor} \}$



与命名实体识别具有同等的形式。

Insight——预测目标转变

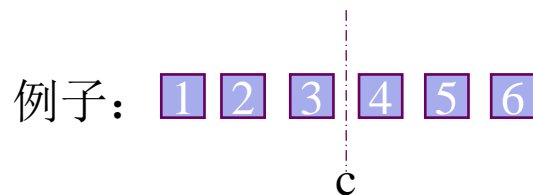
1. 基于边界枚举的方法，存在**错误传播的问题**，当边界预测错误时，分类也会发生错误。
2. 使用两段式的方法，效果好，但是其带来的问题是第一阶段枚举**消耗的时间和空间复杂度为 $O(n^2)$** ；且**后处理复杂**，不利于实际应用。
3. 使用BIEO的识别方法，**只能引入类别信息，而无法有效的引入边界信息**。



中心度 (Centerness): (s_1, e_1) 这么一个片段当成一个特殊单元。其中心 c 为 $c = \frac{(s_1 + e_1)}{2}$ 。对于特殊单元中的其余 token，其中心度定义为偏离中心的程度：

$$c_i = 1 - \frac{|p_i - c|}{s_1 - e_1 + 1}$$

$c_i \in (0.5, 1]$, 越接近 0.5 越可能是边界词
越接近 1, 越可能是中间词。



位置 6 的中心度为: $c_6 = 1 - \frac{6 - 3.5}{6} = \frac{7}{12}$ \longrightarrow 高概率是边界词

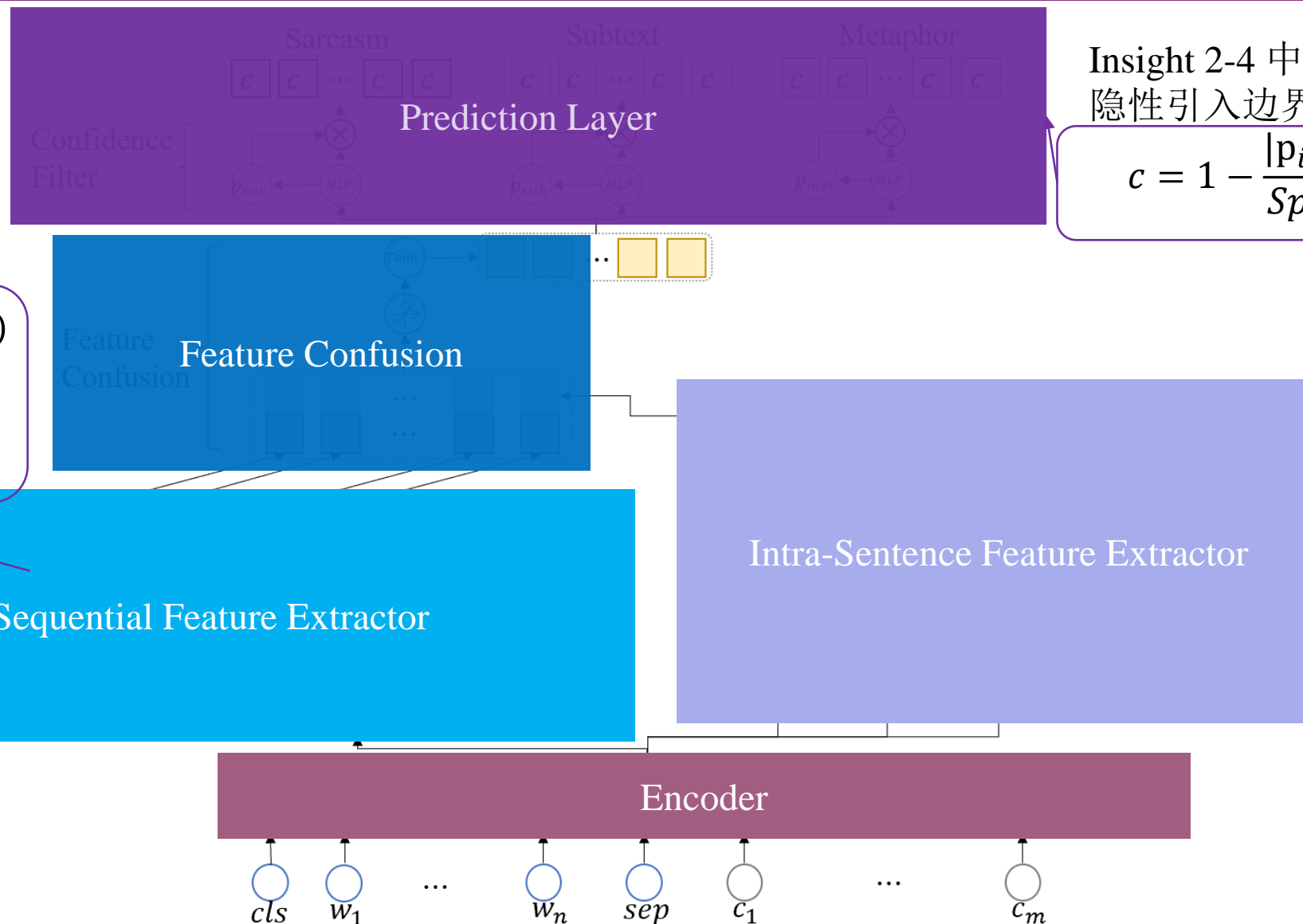
位置 3 和 4 的中心度为: $c_3 = 1 - \frac{0.5}{6} = \frac{11}{12}$ \longrightarrow 中间词

Insight 5. 上下文深度特征建模

$$F_{DFML} = MLP([F_{min}; F_{max}])$$

$$F_{min} = \min(\vec{f}_i, \overleftarrow{f}_i)$$

$$F_{max} = \max(\vec{f}_i, \overleftarrow{f}_i)$$



Insight 2-4 中心度建模，
隐性引入边界与类别信息。

$$c = 1 - \frac{|p_i - center|}{SpanLength}$$



表 5-2: 不同模型在潜台词片段识别上的 IoU 评估结果

Model	IoU_{subt}	IoU_{sarc}	IoU_{meta}	Inference Time
CBM	0.5690	0.4897	0.3467	40 s
BLSM	0.5222	0.4421	0.3088	37 s
BLSM*	0.5325	0.4482	0.3012	38 s
BENSC*	0.5198	0.4311	0.3010	41 s
MHSA	0.5198	0.4524	0.3175	65 s
MHSA*	0.5245	0.4535	0.3117	75 s
LAL	0.5485	0.4551	0.3378	125 s

表 5-3: 不同模型在潜台词片段识别上的 F_1 评估结果

Model	$p_{subt}(\%)$	$r_{subt}(\%)$	$F_{1,subt}(\%)$	$p_{meta}(\%)$	$r_{meta}(\%)$	$F_{1,meta}(\%)$	$p_{sarc}(\%)$	$r_{sarc}(\%)$	$F_{1,sarc}(\%)$	Macro	$F_1(\%)$
CBM	73.44	41.59	53.11	56.83	16.36	25.40	87.83	15.81	26.79		36.71
BLSM	68.13	39.49	50.00	58.27	13.80	22.37	84.35	15.25	25.83		33.84
BLSM*	68.75	39.71	50.34	58.99	13.80	22.37	84.35	15.25	25.83		33.84
BENSC*	70.75	38.71	50.04	58.27	13.59	22.04	84.35	15.25	25.83		33.73
MHSA	69.53	39.71	50.55	59.71	13.97	22.65	84.93	15.55	26.28		33.93
MHSA*	68.85	39.78	50.43	61.15	14.29	23.16	84.93	15.55	26.28		34.08
LAL	70.95	40.62	51.66	61.15	14.39	23.30	84.38	15.25	25.83		34.34

1. **CBM** 速度快，效果最好
2. 两段式的方法**LAL**效果次之，但是速度较慢。
3. 对普通潜台词的识别效果最好，比喻和反讽类型识别结果较差

表 5-4: 消融实验在 IoU 上的评估结果

Model	IoU_{subt}	IoU_{sarc}	IoU_{meta}
CBM	0.5690	0.4897	0.3467
BCBM	0.5479	0.4773	0.3246
SLBM	0.5532	0.4690	0.3306
CMBM	0.5554	0.4725	0.3240

BCBM: centerness -> 二分类

SLBM: centerness -> 多分类

CMBM: 使用拼接进行特征融合

表 5-5: 消融实验在 F_1 上的评估结果

Model	$p_{subt}(\%)$	$r_{subt}(\%)$	$F_{1,subt}(\%)$	$p_{meta}(\%)$	$r_{meta}(\%)$	$F_{1,meta}(\%)$	$p_{sarc}(\%)$	$r_{sarc}(\%)$	$F_{1,sarc}(\%)$	Macro - $F_1(\%)$
CBM	73.44	41.59	53.11	56.83	16.36	25.40	87.83	15.81	26.79	36.71
BCBM	70.63	40.57	51.54	61.15	14.36	23.26	85.22	15.48	26.20	34.72
SLBM	71.25	40.71	51.82	60.43	14.21	23.01	85.22	15.43	26.13	34.69
CMBM	71.88	40.78	52.04	56.12	13.81	22.16	86.09	15.54	26.33	35.29

1. 中心度作为预测目标, 可以有效提升识别效果
2. 分类作为预测目标效果较差, 多分类效果优于二分类
3. DFML层优于拼接的方法, 且不会引入多余变量和计算代价

中心度在嵌套命名实体上也有重要作用!



实际应用

The System of Data Collection &
Subtext Analysis
数据搜集与潜台词分析系统



潜台词识别演示系统

目标语句：
你真好，好看的让我吃不下饭

相关上下文语句：
我好看吗

普通潜台词：
1(有)

反讽：
1(有)

比喻：
-1(无)

夸张：
0(不确定)

谐音：
-1(无)

情绪：
Select a label

态度：
None(无)
生气(Anger)
害怕(Fear)
信任(Trust)
厌恶(Disgust)
难过(Sad) 厌恶(Disgust)
开心(Joy)
惊讶(Surprise)

取消 提交

潜台词识别演示系统

刷新 下载

comment	context	subtext	sarcasm	metaphor	exaggeration	emotion	sentiment
你真好，好看的让我吃不下饭	我好看吗	Yes	Yes	No	Unsure	Disgust	Negative
那你别吃了	你真好，好看的让我吃不下饭	No	No	No	No	Anger	Negative
人是铁，饭是钢	那你别吃了	Yes	No	Yes	No	None	Neutral

data.json 打开文件 查看更多

< 1 >



潜台词识别演示系统

Data Collection

Data Display

Subtext Analysis

需要分析的目标语句：

后来她从深渊里爬出来了却好像死在了那个深渊里

目标语句的完整上下文：

她在婚姻的深渊里苦苦挣扎



取消

提交预测

潜台词判断识别结果

潜台词 6% -1 4% 0 88% +1

无任何反讽 94% -1 1% 0 3% +1

比喻 29% -1 10% 0 59% +1

潜台词内容识别结果

后来她 **从深渊里爬出来了** 却好像死在了那个深渊

后来她从深渊里爬出来了却好像死在了那个深渊

后来她从深 **渊里** 爬出来了却好像死在了那个深渊

推荐回答

你用比喻想说明什么?



相关成果

The Achievement

研究生期间工作成果



学术论文

- **Hua Yan**, Feng Han, Junyi An, Weikang Xiao, Jian Zhao, Furao Shen, “SASICM: A Multi-Task Benchmark For Subtext Recognition,” arXiv preprint arXiv:2106.06944(2021).
- **Hua Yan**, Suhan Guo, Junyi An, Feng Han, Jian Zhao, Furao Shen, “Subtext Recognition: Teach Machine To Read Between The Lines” Coling2022 under review.

专利

- 李俊, **严骅**, 刘晓涛. “一种基于自然语言处理技术的网页文本内容的分类方法” (202110718603.5)
- 李俊, 刘晓涛, **严骅**. “一种基于统计和桩定位视觉的快速检测和定位方法” (202110718272.5)

项目

- 国家自然科学基金项目“基于深度感知增量式联想记忆神经网络的信息融合系统研究”

奖项

- 南京大学英才奖学金
- 南京大学优秀研究生



总结&展望

Conclusion & Future Work

总结与工作展望



中文潜台词数据集与TAE score

- 中文潜台词数据集的构建
- TAE score——新的衡量方法

句子级潜台词分析

- Strengthen Attention
- Multi-Task model(SASICM)

片段级潜台词分析

- 中心度
- 深度特征融合模块
- 嵌套NER上的应用

数据搜集与潜台词分析系统

- 数据搜集与展示功能
- 句子级与片段级潜台词分析功能



数据集部分

- 扩充正式化文本的潜台词数据
- 针对性扩充正样本数据

潜台词检测部分

- 对Strengthen Attention的优化
- 对潜台词特征的提取加强解释性建模
- 上下文语句之间更合理的建模方式

潜台词片段识别

- 中心度的建模通过语义中心进行建模
- 使用非线性方式进行中心度建模

谢谢！请批评指正！

机器人智能与神经计算研究组 RINC
Robotic Intelligence & Neural Computing Group

