Beijing Forest Studio
北京理工大学信息系统及安全对抗实验中心

论辩领域观点对识别以及抽取方法

硕士研究生 高依萌

2022年06月19日

内容提要

- 背景简介
- 基本概念
- 算法原理
- 优劣分析
- 应用总结
- 参考文献

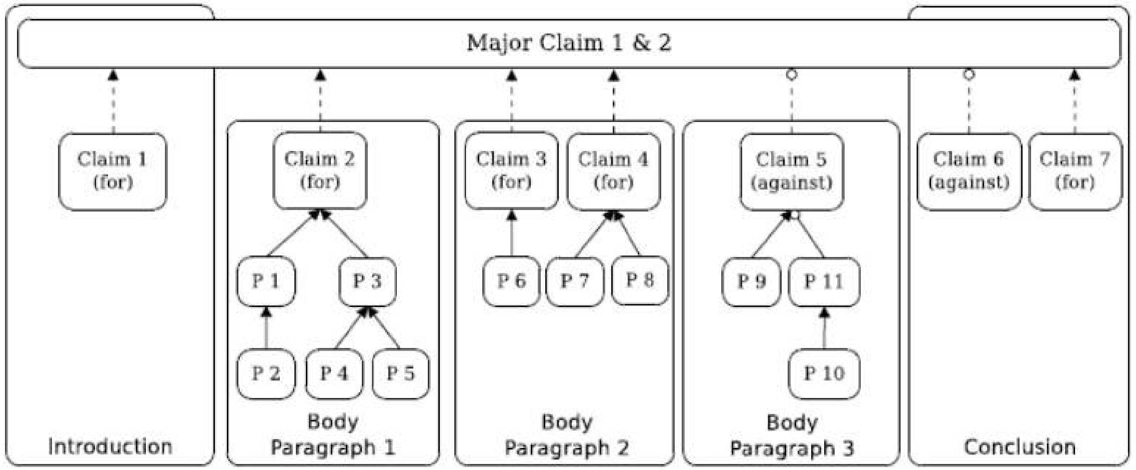
预期收获

- 预期收获
  - 1.了解论辩挖掘领域的基本任务以及经典的系统处理流程
  - 2.了解论辩挖掘领域知识图构建方法以及应用
  - 3.掌握观点对识别和抽取任务定义以及实现方法
  - 4.思考论辩挖掘领域的实际应用和发展方向

背景简介

- 议论文的特点?
  - 议论文由观点组成，观点间的结构和关系表达了文章的内部结构和逻辑关系
- 如何评估议论文的质量?
  - 逻辑论证的过程
  - 老师->机器

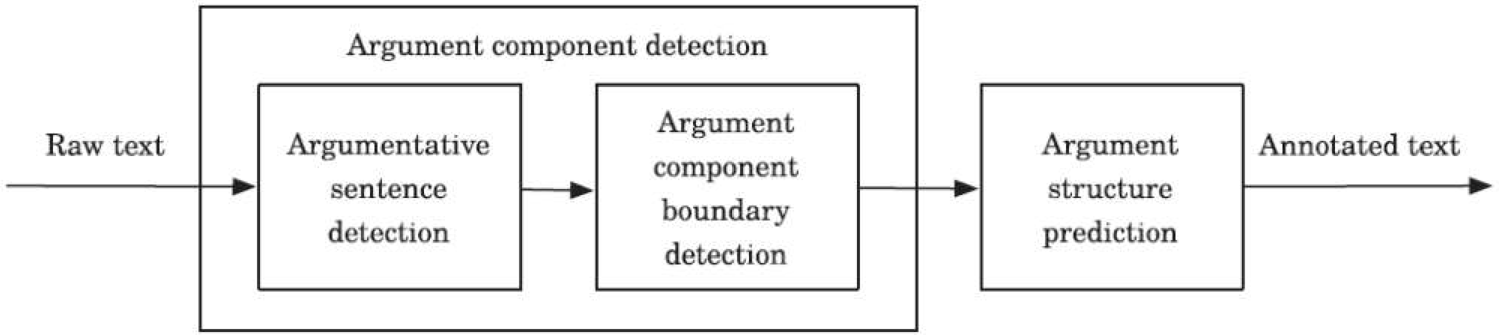
论辩挖掘领域



议论文中论辩结构示意图

背景简介

- 论辩挖掘的基本任务（Argumentation Mining）
  - 从非结构化的论辩性文本中自动抽取论辩文本中的论点，分析论点的内部结构以及不同论点间的关系，最终提供结构化的论辩知识。
    - 第一阶段：论辩元素检测
      - 论辩句子检测，论点部件边界检测（前提，主张） （证据） （结论）
    - 第二阶段：论辩结构预测（观点间的关系）

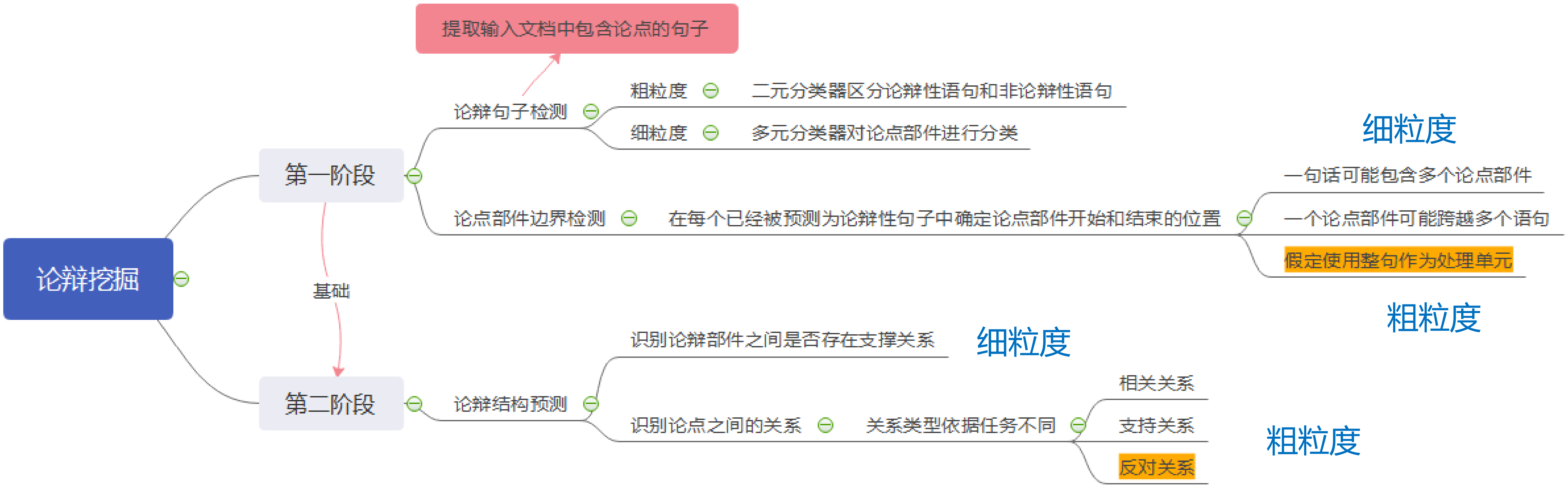


论辩句子检测　论点部件边界检测　论辩结构预测

典型论辩挖掘系统的处理流程

背景简介

- 论辩挖掘任务分类



基本概念

基本概念

- 什么是观点对?
  - 对同一个主题表明立场，支持或者反对。
  - 观点一和观点二就是一组具有反对关系的观点对

主题	应该发展核能
观点一	核能减少了温室气体排放。（支持发展核能）
观点二	核能不是可再生资源，对环保不好。（反对发展核能）

讨论同一个主题 → 观点中的实体概念是有联系的

如何建模实体概念的关系?

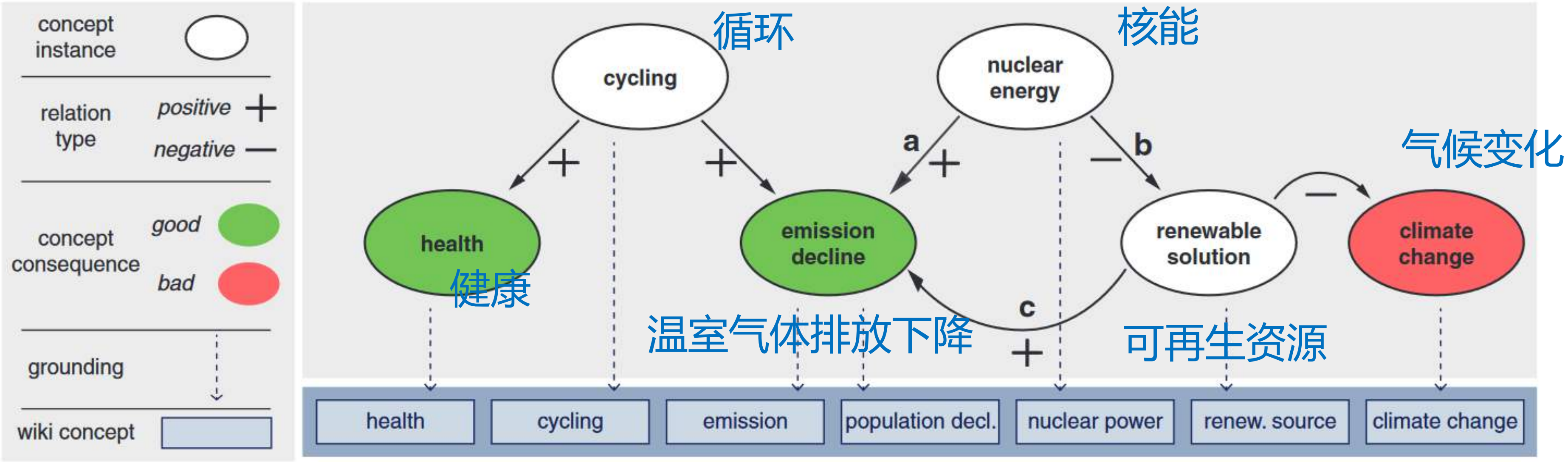
以文本为基础，围绕实体概念构建知识图

https://www.reddit.com/r/changemyview 辩论和交换意见的在线论坛 CMV数据集

基本概念

- 基于Wikipedia知识库建论证知识图
  - Nuclear energy leads to emission decline.
  - Nuclear energy undermines renewable solutions.
  - Renewable solutions tackle climate change and help to decline emission.

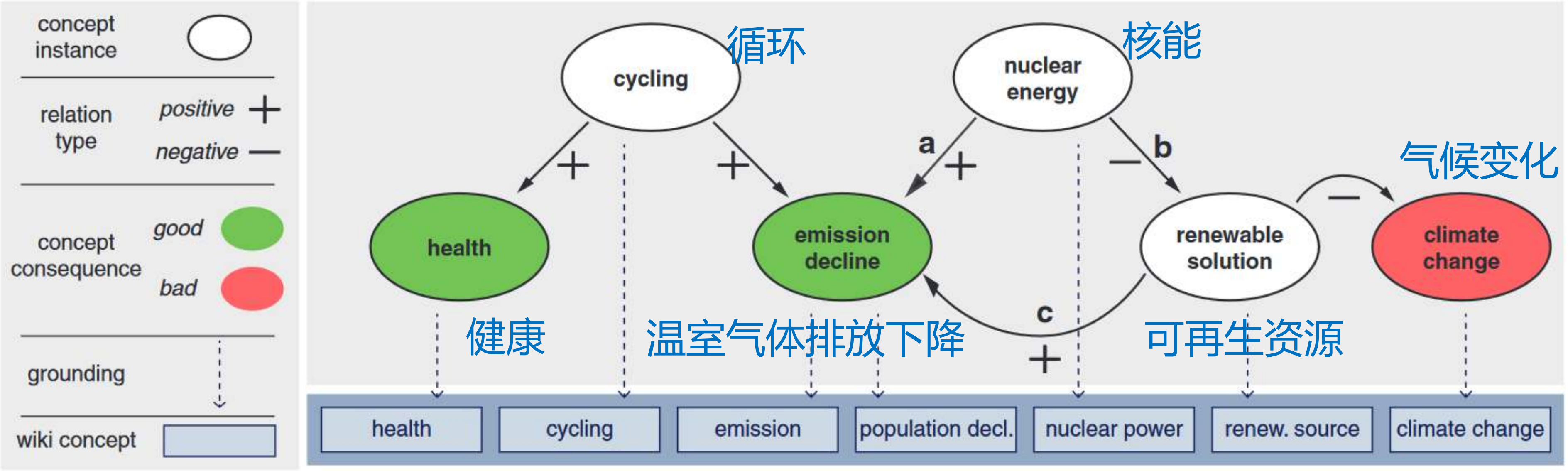
节点（concept instance）是基于Wikipedia的概念构建的



边（relation type）节点间的关系（正向影响，负向影响，或是具体的关系类型）

基本概念

- 论证知识图作用
  - 对于主题“nuclear energy is good for emission decline”
    - 通过（a）就可以支持这个主题
    - 通过（b）（c）这条路径反驳这个主题



基本概念

- 以上例子观点由单句组成
- 观点由多句组成在数据集中如何表示?
  - 由IOBES体系标记观点中的句子
  - O代表不是观点，B是观点的第一句，I是观点的中间句子…
    - The authors propose in this paper to complement the node attributes in a …using a graph embedding algorithm. O O
    - The authors are mixing references to graph convolution for fixed graph ….about general graph neural networks. B-Review B-Review-1
    - We are not sure if we understood this comment well. B-Reply B-Reply-1
    - But, the spatial ….used in all the scenarios is similar. I-Reply I-Reply-1

算法原理

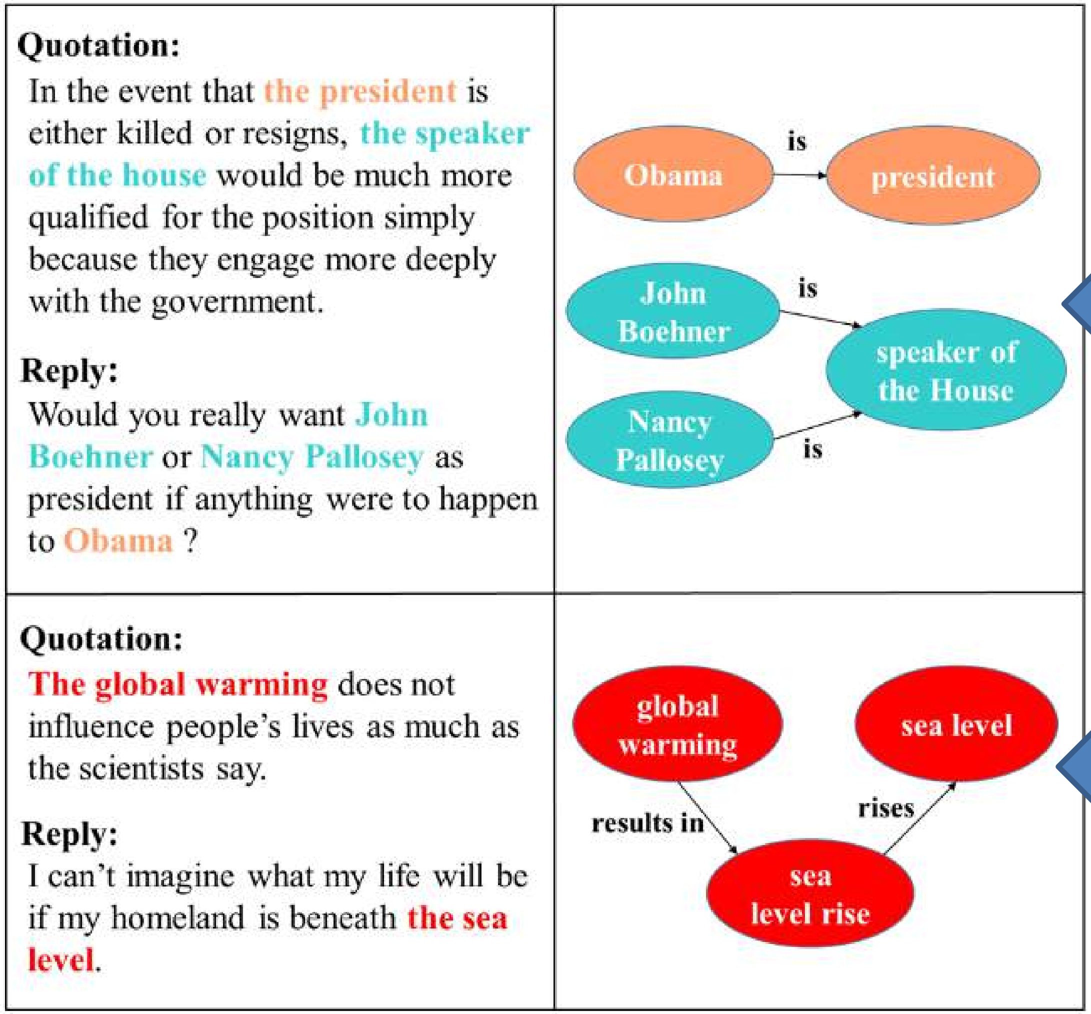
算法原理-LAKG

算法原理-LAKG

T	识别观点对
I	一个观点句q以及其所在文本c_q， 五个候选观点句$\{r_i\}_{i=1}^{5}$以及五个候选观点所在文本$\{c_i\}_{i=1}^{5}$
P	1. 基于CMV数据集构建论证知识图 2. 将观点实体在论证知识图中节点嵌入表示作为背景知识 3. 用transformer编码实体的连接路径作为推理知识 4.整合实体嵌入，路径表示，和观点对文本表示通过信息对齐网络去学习观点对的最终表示以及匹配分数
O	匹配程度最高的观点对（q和r_i）

P	以前的方法将观点对识别看做句子对匹配任务，很大程度上依赖文本描述的相似度，效果不佳
C	观点是整句
D	如何建模观点对内容中的常识知识和推理知识
L	CCF A类会议（ACL 2021）

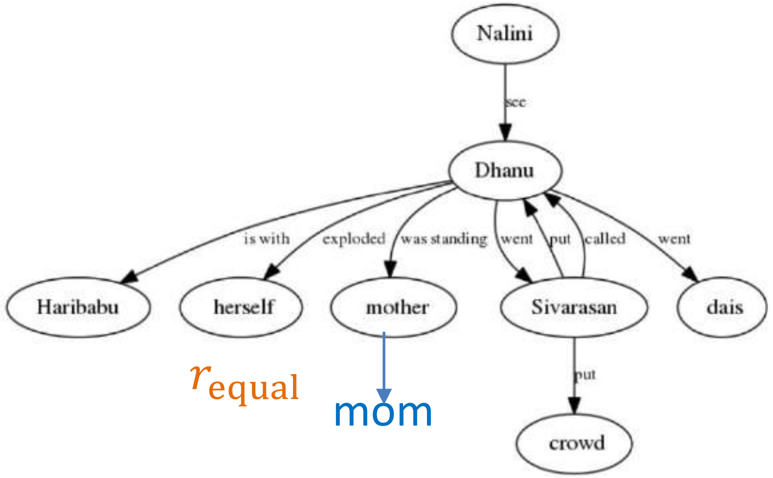
观点对的特点



- 常识知识
  - Obama是president
  - J，N是speaker of the House
- 推理知识（因果关系）
  - 全球变暖会导致海平面上升

算法原理

- 基于CMV数据集构建论证知识图 Stanford OpenIE https://nlp.stanford.edu/software/openie.html
  - 抽取三元组
    - 利用OpenIE抽取第i条数据样本中concept-relation-concept三元组（e_i^h, r_i, e_i^t）
  - Concept Grounding
    - 目的：防止论证知识图稀疏
    - WordNet 词网　Wikipedia API tag me（实体链接工具）
      - 若两节点是同义词或参考了Wikipedia上的同一条目，将两节点间连一条边r_{equal}



Statistics	w/o. grounding	w. grounding
# of nodes	291,199	291,199 --
# of edges	785,036	859,534
avg. degree	2.696	2.952
# of connected components	13,805	10,035

连通分量

Concept grounding操作前后论证图的基础统计

算法原理

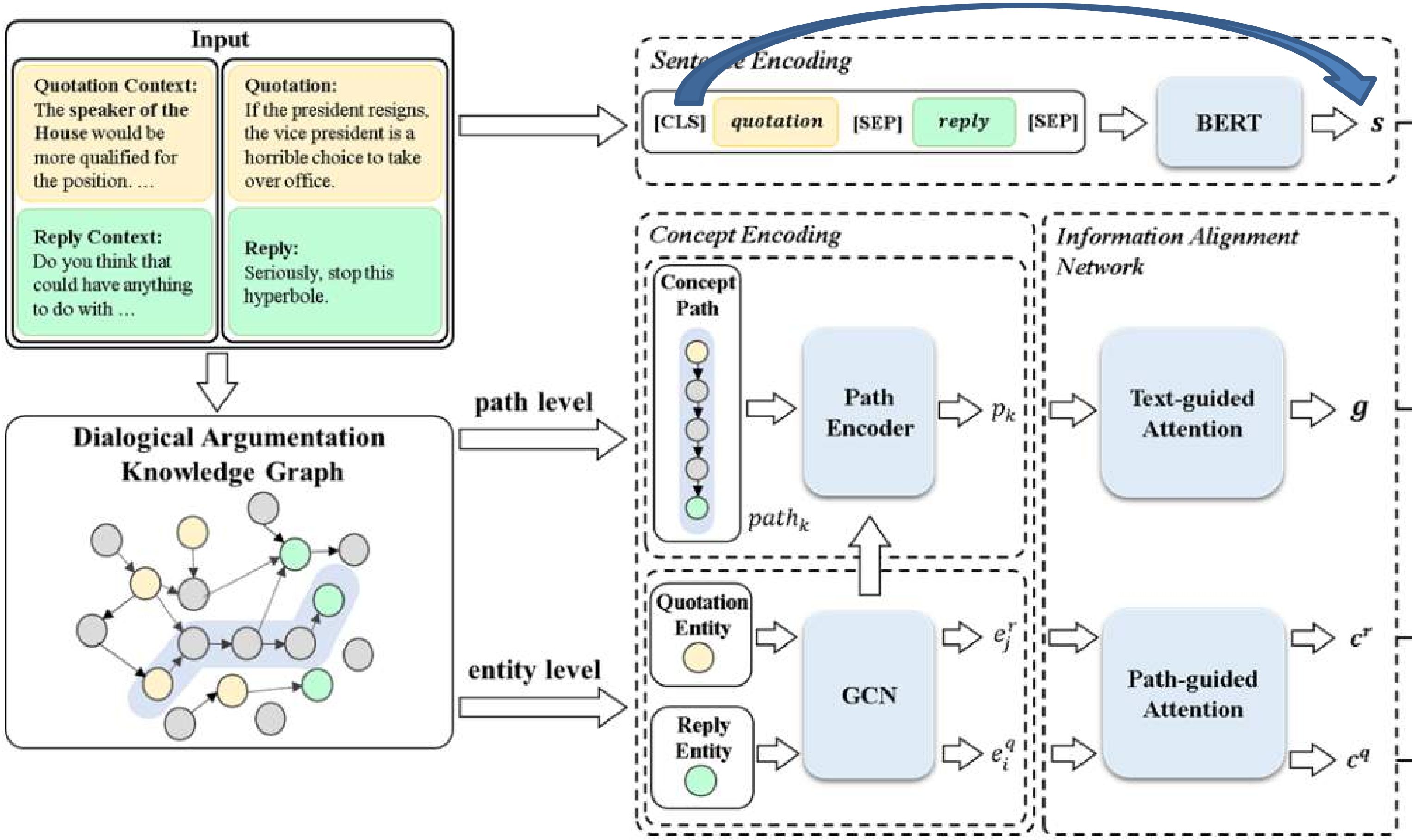


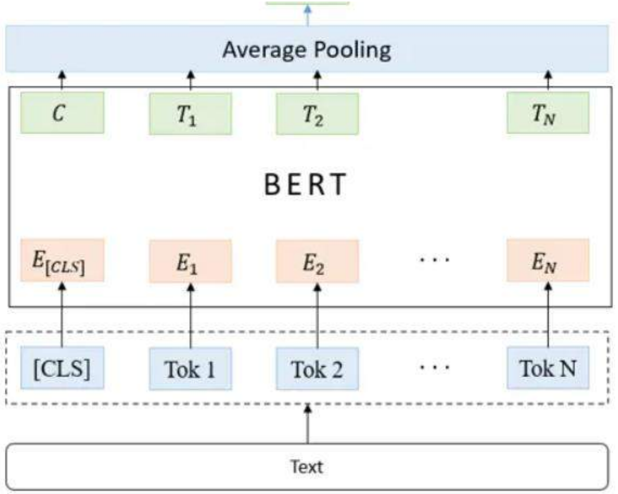
Figure 2: Illustration of the detailed architecture of our model to generate the matching feature vector, which mainly consists of three modules, a Sentence Encoder, a Concept Encoder and an Information Alignment Network. The output of these modules is then fed to a 2-layer perceptron to achieve the final matching score for the given argument pair.

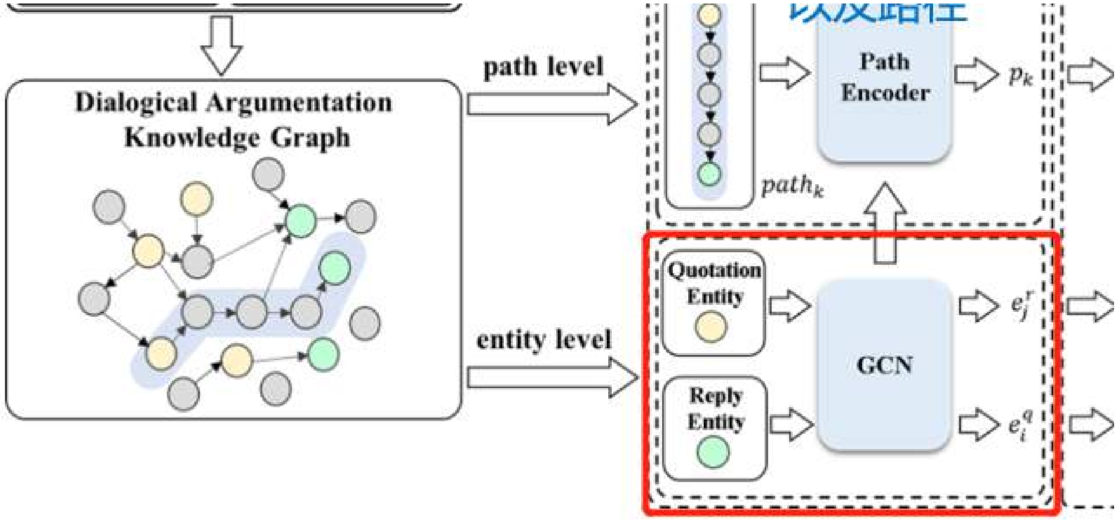
1.如何编码论证图中的节点特征和路径特征?

2.如何整合特征?

算法原理

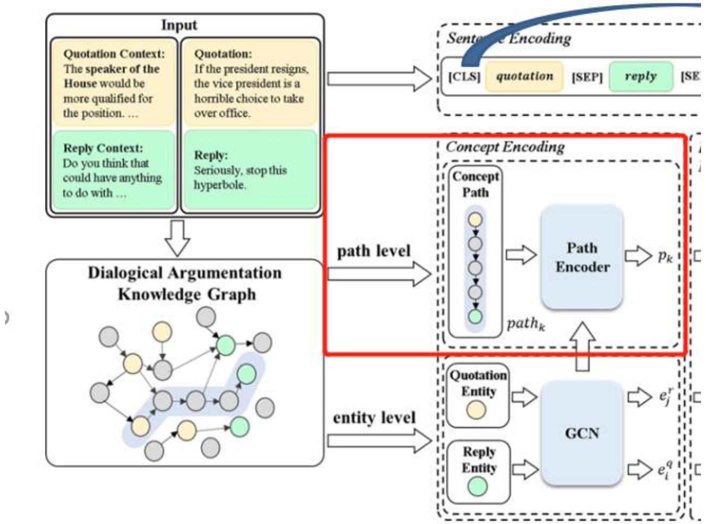
- 实体编码 Concept Encoding
  - 初始化节点表示
    - BERT最后一层作average pooling
  - 用GCN获取实体概念节点表示（entity level）
    - 2层GCN汇聚节点的邻居节点特征
      - $\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$
      - $L = \sigma(\tilde{A}\sigma(\tilde{A}XW_0)W_1)$
        - » $X \in R^{n \times d}$代表n个节点的嵌入矩阵
        - » d代表每个节点的嵌入维度
        - » $D \in R^{n \times n}$对角度矩阵
        - » $A \in R^{n \times n}$图G的邻接矩阵
        - » $L \in R^{n \times n}$图表示
        - » σ为$RELU$函数





算法原理

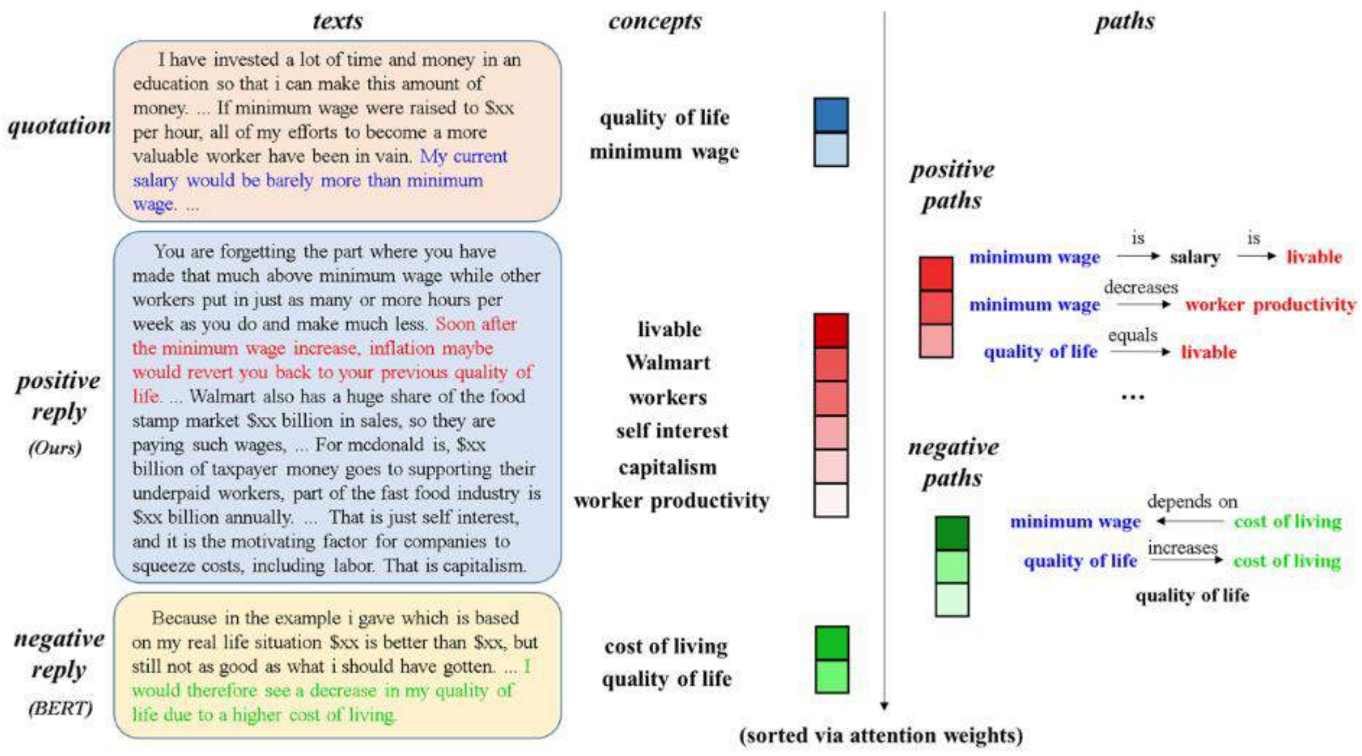
- 路径级别表示 Path Level Representation
  - 编码知识图中路径特征
    - 路径：quotation context的实体节点->…邻居节点…->reply context的实体节点
    - 若有多条路径，选最短路径
    - 路径看作一条序列->使用transformer进行路径编码，并融入位置特征
      - $p_{ij} = Transformer_{Encoder(P_{ij}+PE)}$
      - $P_{ij} = \left(c_i^q, c_1, c_{m_{ij}-1}, c_j^r\right) \in R^{m_{ij} \times d}$
      - P_{ij}位置特征就是节点位于路径中的第j个位置
      - m_{ij}是路径P_{ij}的长度



算法原理

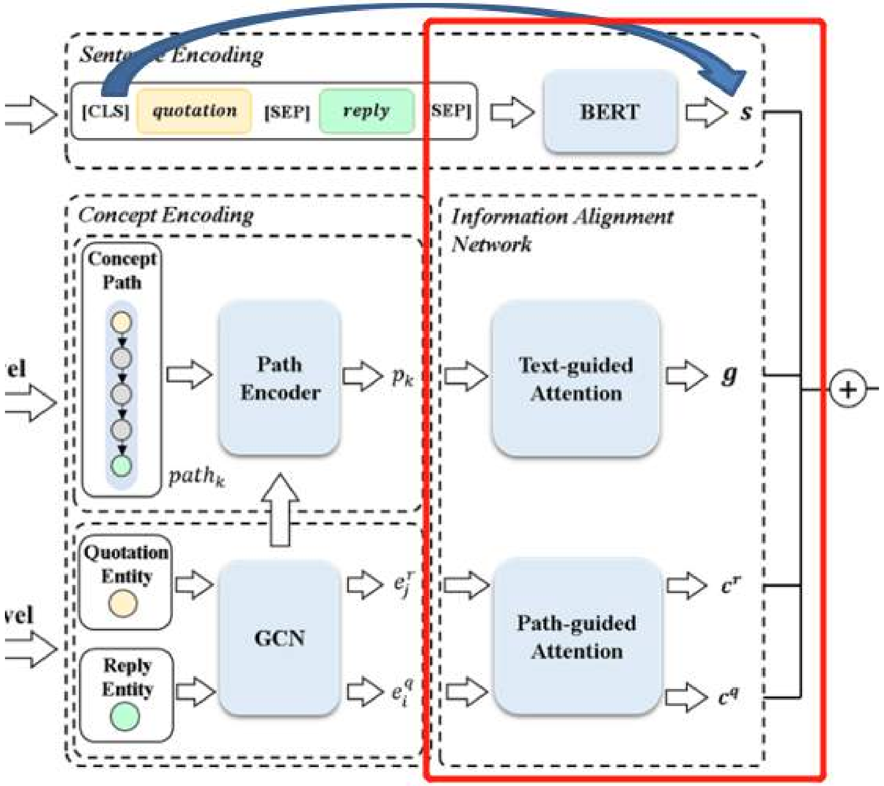
- 信息对齐网络 Information Alignment Network
  - 利用Sentence encoding模块输出的语义匹配特征s整合多条路径特征g
    - 与s匹配程度大的路径赋予高权重，匹配程度小的路径赋予低权重，
  - 利用整合后的路径特征g，为路径中的实体节点赋予不同权重

1.观点对语义匹配特征s
2.观点文本不同实体间的路径
3.每条路径上有n个实体(n>=2)



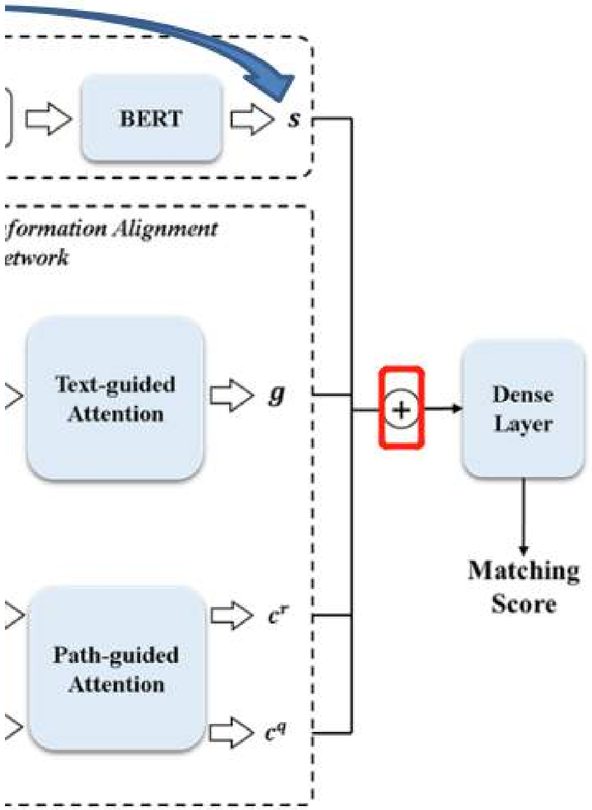
算法原理

- 信息对齐网络 Information Alignment Network
  - Text-guided Attention over Paths
    - $\alpha_k = s(W_2 p_k)$
    - $\hat{\alpha} = SoftMax(\alpha)$
    - $g = \sum_k \widehat{\alpha_k} p_k$
    - p_k为第k条路径，g为整合了所有路径特征的结果
  - Path-guided Attention over Concepts
    - $\beta_i^s = g W_3^s e_i^s$
    - $\hat{\beta}^s = SoftMax(\beta^s)$
    - $c^s = \sum_i \hat{\beta}_i e_i^s$
    - $s \epsilon \{q, r\}$代表概念来自quotation或者reply



算法原理

- 观点对匹配分数 Matching Score
  - 将观点对语义匹配特征s，路径特征g，实体特征$c^q c^r$拼接分类
    - $f = [s; g; c^q; c^r]$
    - $S = \sigma(W_s f + b_s)$
  - MarginRankingLoss
    - 目的是让loss趋近于0
    - 预期让正样本对的匹配分数比负样本对的匹配分数大
    - $L = \sum_{i=1}^{4} \max(0, -(S(q, r^+) - S(q, r_i^-)) + \gamma)$
    - γ可变的加在loss上的偏移量，超参数



实验结果

- 对比实验

Methods	P@1(%) Dev Set	P@1(%) Test Set	MRR(%) Dev Set	MRR(%) Test Set
Random Guess	20	20	45.67	45.67
BiGRU	65.92	51.52	75.22	70.57
BiGRU+RNN Context	69.29	55.98	80.51	73.20
BiGRU+Hierarchical Context	70.93	57.46	82.47	73.72
VAE+Hierarchical Context	71.28	58.61	83.82	74.66
DVAE+Hierarchical Context*	73.70	61.17	85.14	76.16
BERT	73.18	61.85	84.69	76.57
BERT+Hierarchical Context	76.81	66.85	86.38	78.51
Ours	78.33	**68.75**	87.43	**80.85**

Table 2: Performance comparison for all the models on the development dataset and the test dataset, where the sign '*' represents the former state-of-the-art model. The best result on the test set is in **bold**

  - BERT在不利用上下文信息的情况下，观点对识别效果基本是最优
  - 利用context信息有利于提升识别观点对的效果
  - 基于论证知识图的方法>层次网络建模上下文的方法>RNN建模上下文

实验结果

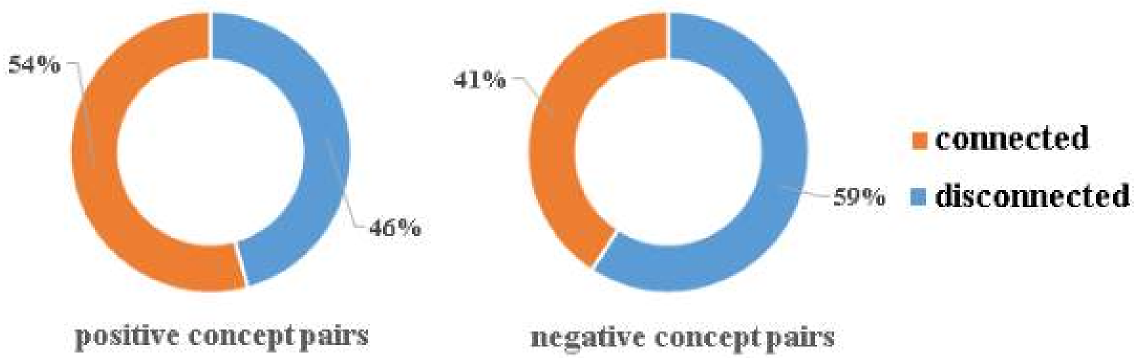
- 消融实验
  - （1）Transformer路径编码效果比BiGRU好
  - （2）去掉路径特征模型效果在P@1中下降4%
  - （3）路径中的实体概念特征对观点对的交互有重要的指导作用
  - （4）模型依赖Bert对观点对的编码特征

	Model	P@1	MRR
	Sentence + Concept + Align	68.75	80.85
（1）	: BiGRU as Path Encoder	67.12	79.46
（1）	: w/o Alignment Layer	65.48	79.39
（1，2）	: w/o Path Transformer	64.41	77.41
（3）	: w/o Concept Encoder	61.85	76.57
（4）	: w/o Sentence Encoder	51.96	68.83

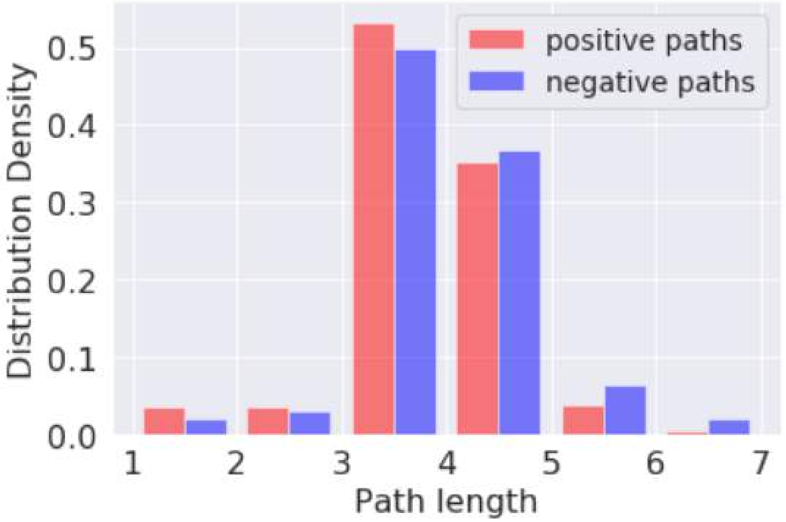
Table 3: Ablation study on our proposed framework.

实验分析

- 正样本对存在连接路径的实体对比负样本对多
  - 说明观点对讨论同一主题，实体间关系更加紧密
- 实体对间的路径长度主要分布在3-4
  - 正样本对中的路径长度比负样本对中的短
  - 说明推理路径越长讨论的内容越偏离主题



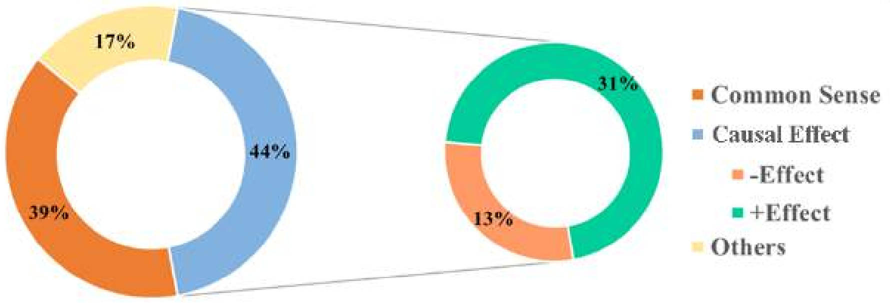
统计观点文本中实体对间路径连接情况



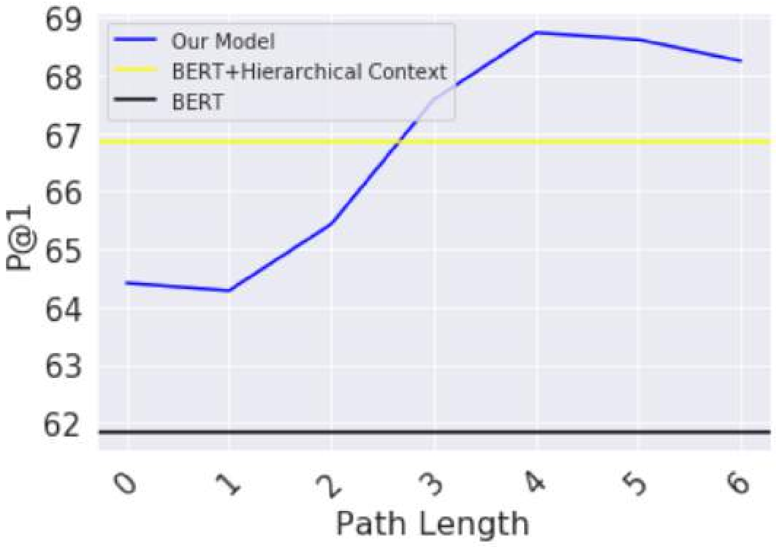
正样本对和负样本对路径长度分布情况

实验分析

- 统计推理路径中常识关系和因果关系分布
  - 常识关系占比近40%，因果关系占比44%
    - 因果关系（正向影响约占比31%，负向影响约占比13%）
  - 路径长度阈值设置为3-4效果最佳



统计所有路径中不同关系分布情况



设置路径长度阈值

优劣分析

- 优势
  - 通过构建知识图显式建模了CMV数据集中隐含的常识知识和推理知识
- 劣势
  - 观点句没有抽取出实体的情况下，无法为观点补充推理路径
  - 当reply短且和quotation没有共现词时，模型识别两者匹配关系效果差
    - [quotation] If the president is either killed or resigns, the vice president is a horrible choice to take over the office.
    - [reply] Seriously, stop this hyperbole.

算法原理

算法原理

算法原理-MGF

算法原理

T	预测两段文本句子的标签，抽取出标签匹配的观点对
I	review和rebuttal两篇文本
P	1.用序列标注器预测两篇文本句子标签(Argument Mining) 2.利用句子间关系图显式建模两段文本内和文本间句子关系 3.将第一步骤中预测的观点的特征分别加到待匹配文本的观点上，预测待匹配文本观点标签（Argument Pair Extraction）
O	带有标签的观点句

P	如何显式建模观点句之间的交互
C	观点由多个句子组成，观点对存在一对多的关系
D	句子间关系图的设计
L	CCF A类会议（ACL 2021）

算法原理

- 观点由多句组成，同颜色为观点对　　　序列标注任务

一个观点

Review	Sent	Arg
This work applies convolutional neural networks to the task of RGB-D indoor scene segmentation.	Sent-1	Non-Arg
…	…	
The model simply adds depth as a separate channel to the existing RGB channels in a conv net.	Sent-3	Rev: Arg-1
Depth has some unique properties e.g. infinity/ missing values depending on the sensor.	Sent-4	
It would be nice to see some consideration or experiments on how to properly integrate depth ···	Sent-5	
The experiments demonstrate that a conv net using depth information is competitive ···	Sent-6	Rev: Arg-2
…	…	
Does this suggest that depth isn't always useful, or that there could be better ways to ···	Sent-9	
…	…	…

Arg-Pair-1

Arg-Pair-2

Arg	Sent	Rebuttal
Non-Arg	Sent-1	Thank you for your review and helpful comments.
Rep: Arg-1	Sent-2	The missing values in the depth acquisition were pre-processed using inpainting code available ···
	Sent-3	We added the reference to the paper.
Rep: Arg-2	Sent-4	In the paper, we made the observation that the classes for which depth fails to outperform the RGB model are the classes of object for which the depth map does not vary too much.
	Sent-5	We now stress out better this observation with the addition of some depth maps at Figure 2.
Non-Arg	…	…
	Sent-7	The current RGBD multiscale network is the best way we found to learn features using depth, ···
	…	…

算法原理

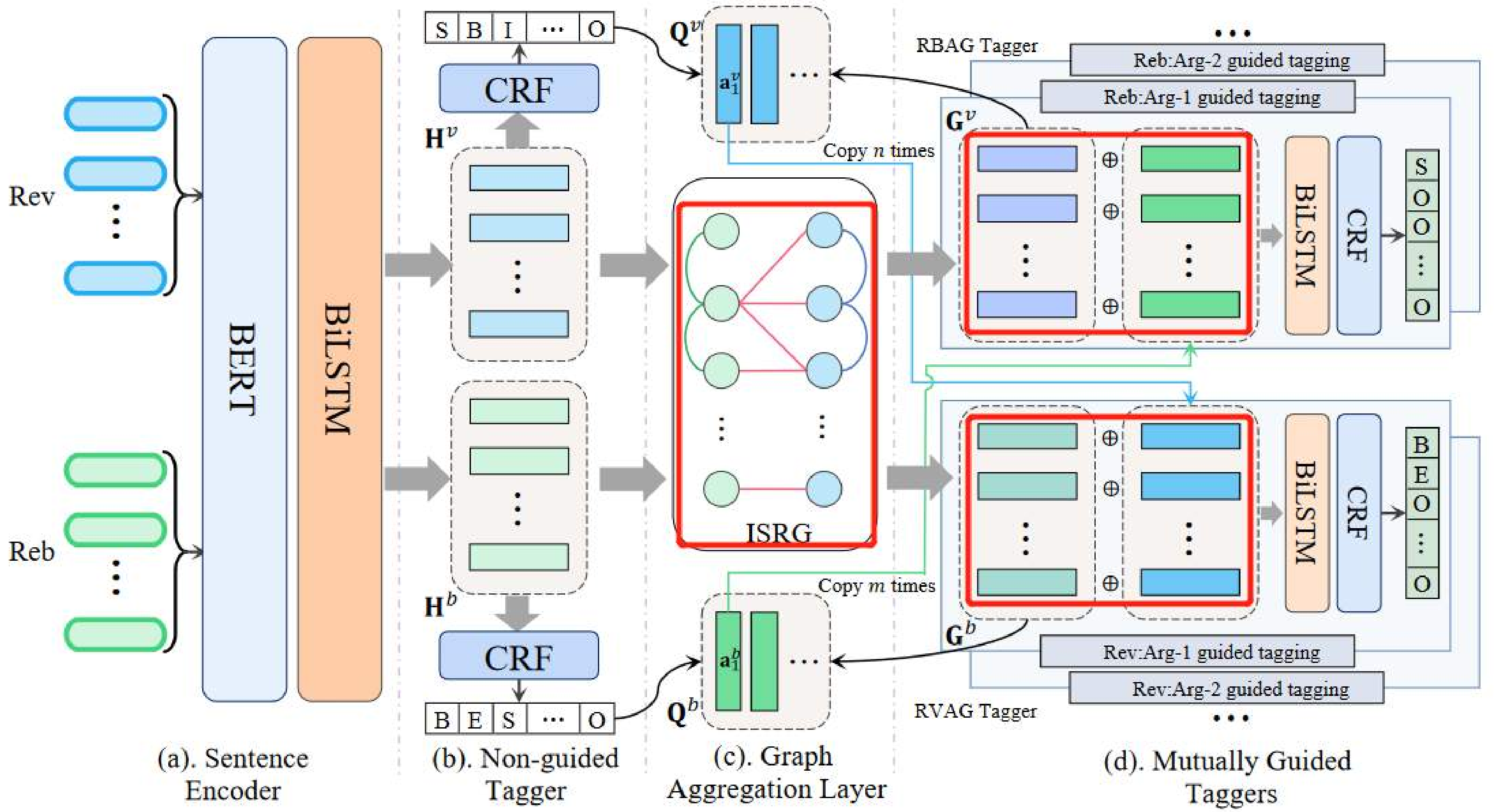


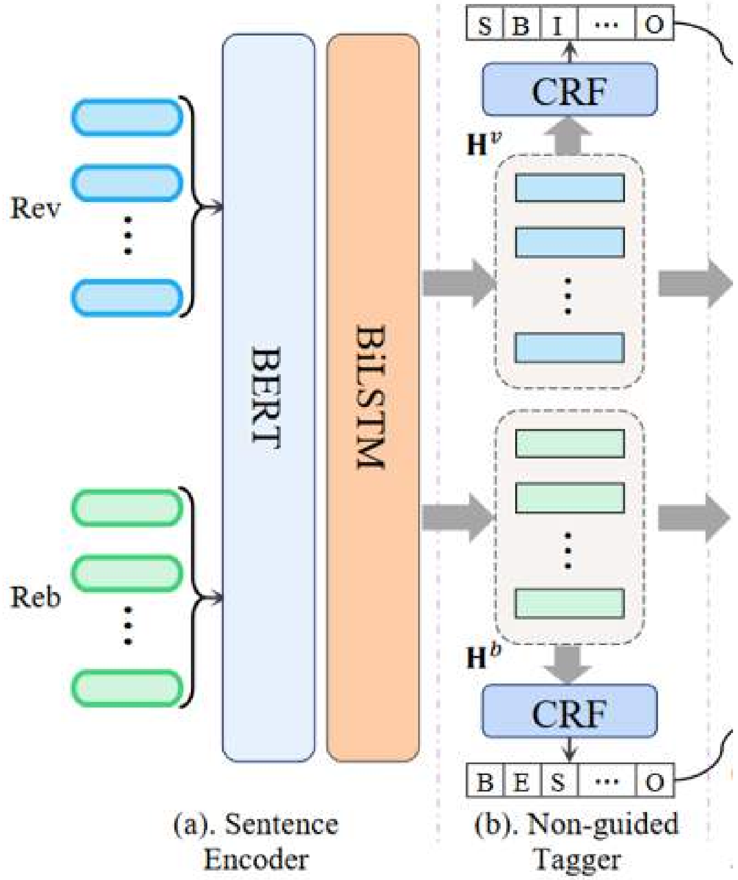
Figure 2: The architecture of MGF.

Argument Mining

Argument Pair Extraction

算法原理

- 相互指导框架 Mutual Guidance Framework
  - Sentence Encoder(BERT+BiLSTM)
    - mean pooling句中所有词表示获得句向量e_i
      - $V = (e_1^v, e_2^v, \cdots e_m^v)$
      - $B = (e_1^b, e_2^b, \cdots e_m^b)$
        - » $e_i \in R^{d_b}$ d_b是BERT最后一层的维度
    - 将句向量e_i送入BiLSTM获取上下文依赖表示
      - $H^v = (h_1^v, h_2^v \cdots h_m^v)$
      - $H^b = \left(h_1^b, h_2^b \cdots h_m^b\right)$
        - » $h_i^v / h_i^v \in R^{2d_l} d_l$是LSTM隐藏层维度
  - Non-guided Tagger
    - 利用CRF给review和rebuttal中每个句子表征H^v和H^b预测句子标签
      - $Y^v = (y_1^v, y_2^v \cdots y_m^v)$
      - $Y^b = \left(y_1^b, y_2^b \cdots y_m^b\right)$



算法原理

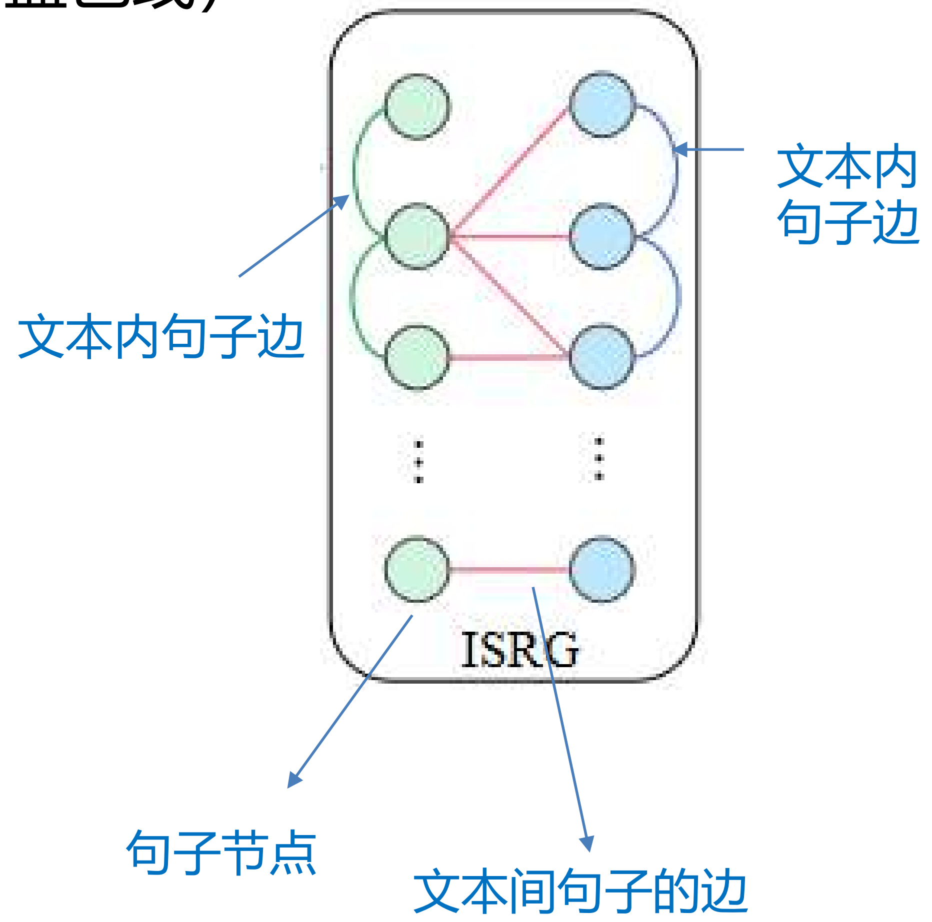
- 句子间关系图 Inter-sentence Relation Graph
  - 观点文本内句子间边的权重如下（绿色线或者蓝色线）

    - $$w^I(s_i, s_j) = \begin{cases} 1 + \left(1 - \frac{D(s_i,s_j)}{\rho}\right) & D(s_i, s_j) \le \rho \\ 0 & otherwise \end{cases}$$

    – ρ是句子距离的阈值
  - 观点文本间句子间边的权重如下（粉色线）

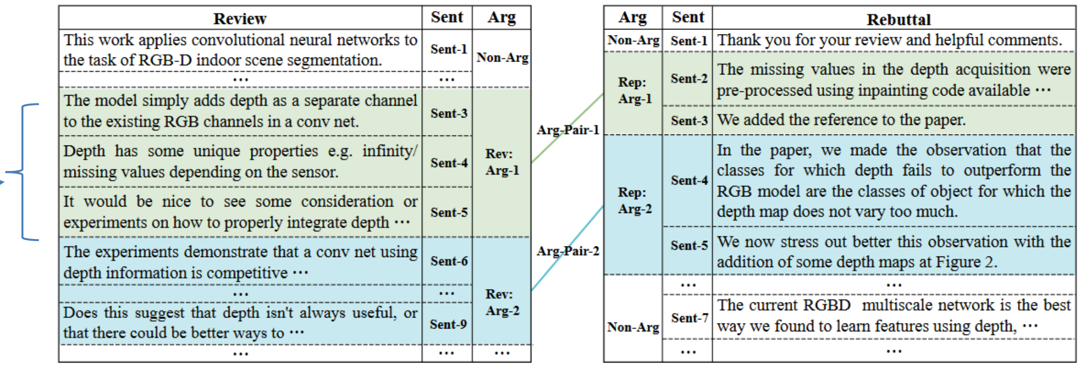
    - $$w^C(s_i, s_j) = \begin{cases} 1 + \frac{C(s_i,s_j)}{C_{max}} & C(s_i, s_j) > \varphi \\ 0 & otherwise \end{cases}$$

    – φ共现词数量的阈值



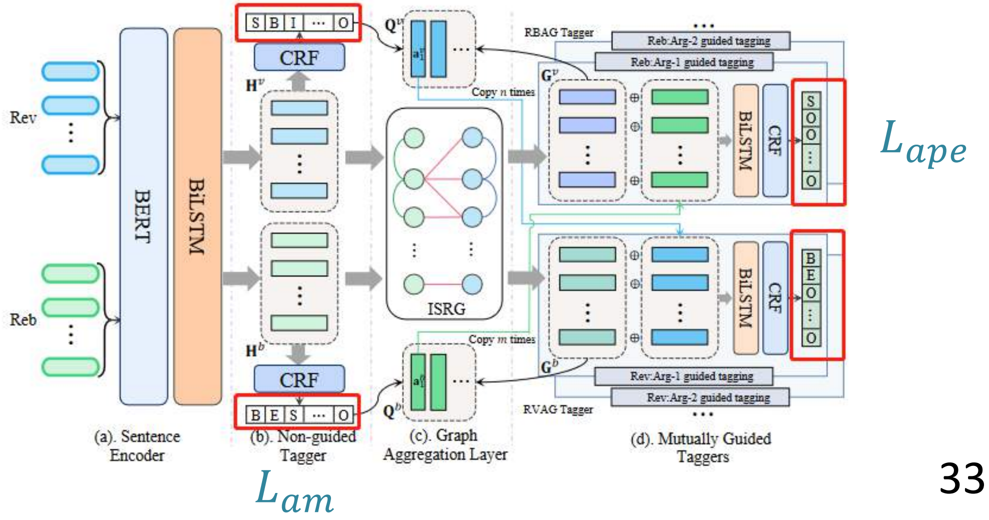
算法原理

- 相互指导标注器 Mutually Guided Taggers
  - RVAG
    - 获取review中的观点表示a_k^v
      - $a_k^v = \frac{1}{e_k - b_k + 1} \sum_{i=b_k}^{e_k} g_i^v$
      - $a_k^v \in (b_k, e_k)$, g_i^v是review的第i个句向量
    - 将a_k^v拼接到rebuttal每个观点向量上，再将拼接结果输入BiLSTM获取最终表征h_i
    - 利用CRF预测h_i的标签，识别其是否与a_k^v能形成观点对
  - RBAG同理
- 损失函数
  - $L = L_{am} + L_{ape}$



Review	Sent	Arg
This work applies convolutional neural networks to the task of RGB-D indoor scene segmentation.	Sent-1	Non-Arg
…	…	
The model simply adds depth as a separate channel to the existing RGB channels in a conv net.	Sent-3	Rev: Arg-1
Depth has some unique properties e.g. infinity/missing values depending on the sensor.	Sent-4	
It would be nice to see some consideration or experiments on how to properly integrate depth ⋯	Sent-5	
The experiments demonstrate that a conv net using depth information is competitive ⋯	Sent-6	Rev: Arg-2
…	…	
Does this suggest that depth isn't always useful, or that there could be better ways to ⋯	Sent-9	
…	…	…

Arg	Sent	Rebuttal
Non-Arg	Sent-1	Thank you for your review and helpful comments.
Rep: Arg-1	Sent-2	The missing values in the depth acquisition were pre-processed using inpainting code available ⋯
	Sent-3	We added the reference to the paper.
Rep: Arg-2	Sent-4	In the paper, we made the observation that the classes for which depth fails to outperform the RGB model are the classes of object for which the depth map does not vary too much.
	Sent-5	We now stress out better this observation with the addition of some depth maps at Figure 2.
	…	…
Non-Arg	Sent-7	The current RGBD multiscale network is the best way we found to learn features using depth, ⋯
	…	…



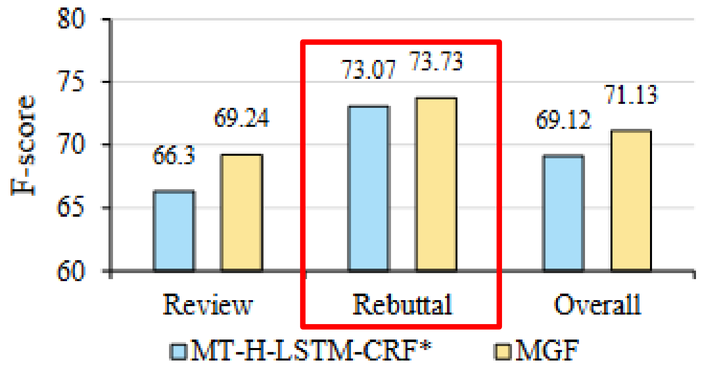
实验结果

- 对比实验
  - RR-submission（观点对是一对多关系）：在F1值上，较当前最优模型MT-H-LSTM-CRF分别在AM和APE任务上提升了至少1.01% 和 7.94%
  - RR-passage （观点对是一对一关系）：在F1值上，较当前最优模型MT-H-LSTM-CRF分别在AM和APE任务上提升了至少0.79%和7.01%

Data	Method	Argument Mining			Argument Pair Extraction		
		Pre.	Rec.	F_1	Pre.	Rec.	F_1
RR-submission	PL-H-LSTM-CRF	67.63	68.51	68.06	19.86	19.94	19.90
	MT-H-LSTM-CRF	70.09	70.14	70.12	26.69	26.24	26.46
	Two-Step	**70.94**	70.77	70.86	33.11	24.67	28.27
	Non-FT-MGF	69.18	69.94	69.55	33.12	33.69	33.40
	MGF (Ours)	70.40	**71.87**	**71.13**	**34.23**	**34.57**	**34.40**
RR-passage	PL-H-LSTM-CRF	73.10	67.65	70.27	21.24	19.30	20.23
	MT-H-LSTM-CRF	71.85	71.01	71.43	30.08	29.55	29.81
	Two-Step	71.94	**71.51**	71.72	34.31	26.87	30.14
	Non-FT-MGF	71.22	70.49	70.85	35.20	34.11	34.65
	MGF (Ours)	**73.62**	70.88	**72.22**	**38.03**	**35.68**	**36.82**

实验结果

- AM对比实验
  - rebuttal实验效果均达到73%以上，rebuttal实验效果只有65%以上
    - 原因：rebuttal比review的结构更清晰
- APE消融实验
  - 句子间关系图模块对模型影响最大，去掉此模块F1值会下降3.75个点
  - 只用RBAG tagger的效果比只用RVAG tagger的F1值好
    - 原因：在AM任务上rebuttal观点的识别比review观点识别更加准确



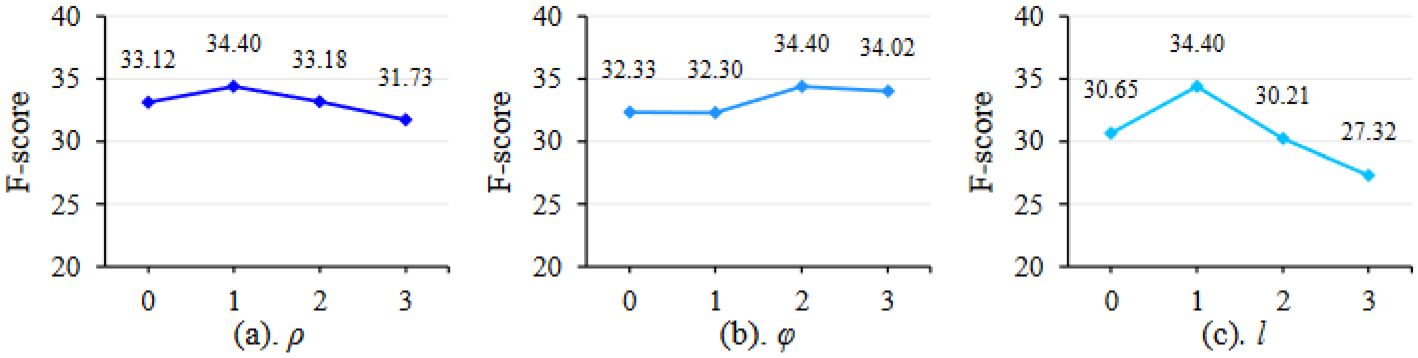
RR-submission数据集上 AM 对比实验结果

Method	APE F_1	∇
MGF (Ours)	**34.40**	-
w/o RVAG Tagger	33.11	-1.29
w/o RBAG Tagger	31.94	-2.46
w/o ISRG	30.65	-3.75
w/o IPE	33.12	-1.28
w/o CPE	32.33	-2.07

消融实验

实验结果

- Impacts of Graph Parameters
  - 句子内距离ρ阈值 1
    - 近距离句子构成观点的可能性更大；远距离的句子可能会扭曲观点语义
  - 共现词φ阈值 2
    - 这表明观点对的共现词为2的概率更大
  - GCN层数l的阈值 1
    - 一层GCN可充分建模句子间关系



优劣分析

- 优势
  - 显式建模了观点文本内和文本间句子的关系
  - 一对一，一对多关系的观点对准确率更高
  - 论辩挖掘领域，对审稿回复类文本进行观点挖掘奠基之作
- 劣势
  - 如果两个观点句间没有共现词，ISRG无法为两个句子节点连接边
  - 论点标识不完整

应用总结

应用总结

应用总结

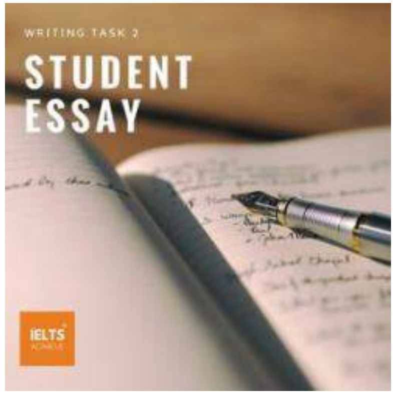
应用总结

- 应用前景广阔
  - 智能法律、医学领域、人文与教育、用户生成内容等领域
  - 自动司法判案、自动医疗诊断和个性化的药物处方、议论文自动评分、争议性观点识别、观点影响力与可信性评估、 观点的检索与呈现
  - 论辩挖掘开源项目：
    - CAIL 2021 论辩理解 http://cail.cipsc.org.cn/task6.html?raceID=4
      https://github.com/Ding-Jiayu/CAIL2021-ArgumentationUnderstanding
    - EMNLP 论辩挖掘研讨会 https://2021.argmining.org/

CAIL 中国法律智能技术评测 Challenge of AI in Law

识别诉辩双方的观点对

建立症状描述和疾病间关联



议论文评分

未来展望

- 数据特点角度
  - 现有研究缺少对多个论点间论辩结构的挖掘
    - 如科学论文全文的论辩结构、社交媒体中原贴和评论组成的多个文档间的论辩结构研究较少
- 技术角度
  - 论辩结构识别主要以有监督学习方法为主
    - 论辩结构复杂，标准制定有争议，语料匮乏
    - 元学习，小样本学习解决语料匮乏的困境

数据集名称	数据集规模
SMP-CAIL2020	3263
CMV	13064
RR	4764
BWS	3310
UKP ASPECT	3595

自主构建小规模数据集，多数限定在特定领域

- 应用角度
  - 主要针对社交媒体中的短文本，科学论文等复杂长文本的研究还相对较少
    - 新闻评论、科学论文、法律判决书等长文本中往往含有价值更高的论辩信息

参考文献

[1] Yuan J, Wei Z, Zhao D, et al. Leveraging Argumentation Knowledge Graph for Interactive Argument Pair Identification[C]//Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. 2021: 2310-2319.

[2] Bao J, Liang B, Sun J, et al. Argument Pair Extraction with Mutual Guidance and Inter-sentence Relation Graph[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021: 3923-3934.

[3] Al-Khatib K, Hou Y, Wachsmuth H, et al. End-to-end argumentation knowledge graph construction[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(05): 7367-7374.

[4]宋巍 1，魏忠钰 2 1首都师范大学，2复旦大学 论辩挖掘研究

[5]李永泽, 欧石燕. 论辩挖掘研究综述[J]. 图书情报工作, 2020, 64(19): 128.

谢谢！

大成若缺，其用不弊。大盈若冲，其用不穷。大直若屈。大巧若拙。大辩若讷。静胜躁，寒胜热。清静为天下正。