

文章编号: 1003-0077(2023)10-0108-14

对话式论辩研究综述

魏忠钰¹, 丁佳琦¹, 沈晨晨¹, 高 源¹, 梁敬聪¹, 纪程炜¹, 林嘉昱¹, 黄萱菁²

(1. 复旦大学 大数据学院, 上海 200433;

2. 复旦大学 计算机科学技术学院, 上海 200433)

摘 要: 近年来, 论辩研究引起计算语言学学者的关注, 并催生了一个新的研究领域, 即计算论辩学。根据参与论辩过程的人数不同, 计算论辩学的研究可以分成两类, 即, 单体式论辩和对话式论辩。对话式论辩过程在现实世界中广泛存在, 如社交网络平台、司法领域、教育领域等, 但是相关的研究才刚刚起步。该文综述了对话式论辩领域的基本任务设置、主流模型框架、下游应用以及公开数据和评测方法。最后, 该文也指出对话式论辩未来发展的几个研究方向, 包括多模态的对话式论辩分析、知识注入的论辩生成等。

关键词: 计算论辩; 对话式论辩

中图分类号: TP391

文献标识码: A

A Survey of Dialogical Argumentation

WEI Zhongyu¹, DING Jiayu¹, SHEN Chenchen¹, GAO Yuan¹, LIANG Jingcong¹,

JI Chengwei¹, LIN Jiayu¹, HUANG Xuanjing²

(1. School of Data Science, Fudan University, Shanghai 200433, China;

2. School of Computer Science, Fudan University, Shanghai 200433, China)

Abstract: In recent years, the study of argumentation has given birth to a new research field, i.e. computational argumentation. According to the number of people participating in the argumentation process, the research of computational argumentation can be divided into two categories, namely, monological argumentation and dialogical argumentation. The process of dialogical argumentation happens widely in the real world, such as social network platforms, judicial fields, education fields, etc., but related research has just started. This paper reviews the basic task settings, mainstream model frameworks, downstream applications, and public datasets and evaluation methods in the field of dialogical argumentation. Finally, this paper reveals several research directions for the future development, including multi-modal dialogue argumentation analysis, knowledge-infused argumentation generation, etc.

Keywords: computational argumentation; dialogical argumentation

0 引言

论辩(Argumentation)^[1-3]以人的逻辑论证过程作为研究对象, 是一个涉及逻辑、哲学、语言、修辞、计算机科学和教育等多学科的研究领域, 相关研究工作可以追溯到亚里士多德时期^[4]。近年来, 论辩研究引起计算语言学的关注, 并催生了一个新的研究领域, 即计算论辩学(Computational Argumentation)^[5-6]。学者试图将人类关于逻辑论证的认知模型与计算模型结合起来, 以提高人工智能自动推理的能力^[7-8]。

根据参与论辩过程的人数不同, 计算论辩学的研究可以分成两类, 即, 单体式论辩(Monological Argumentation)和对话式论辩(Dialogical Argumentation)。单体式论辩^[6]的研究对象是仅有一个参与者的辩论性文本, 如议论文和主题演讲等。相关的研究包括论辩部件检测^[9]、论辩结构预测^[10-11]、论辩策略分类^[12-13]和议论文评分^[14-15]等。对话式论辩^[16-17]的研究对象是针对某一个特定议题进行观点交互的论辩过程, 一般有多个参与者。相关的研究包括论辩结果预测^[18-19]和论辩逻辑链抽取^[20-21]等。近年来, 相关学者在单体式论辩研究中取得了

一定程度的进展,包括任务类型的确定、数据集合的构建以及基线模型的提出等。但对于对话式论辩的相关研究还停留在初步阶段,缺乏对相关场景和任务的系统化梳理与建模。

多参与者的论辩交互过程在现实世界中广泛存在,如社交网络^[22-23]、司法领域^[24-26]、教育领域^[27-30]等。随着社交媒体(微博、微信、知乎、脸书等)在世界范围内的影响力不断扩大,人们交流信息的方式也发生了革命性的变化。很多话题的讨论都从“餐桌对话”的形式扩展到在线平台,大量积累的用户讨论文本也因此成为大家对公众议题进行搜索的重要信息来源^[31]。在法庭的案件审理过程中,控诉方和辩护方会针对一个特定案件的不同侧面进行观点的交换,而法官和陪审团则会基于这样的交互信息,对案件的结果进行最终的判决。每一法庭庭审记录都包含大量控辩双方的交互文本,成为法官和陪审团解读案件需要进行深度理解的材料。在教育领域,议论文写作是学生在语言课程学习中的必修科目,学生需要对于给定主题给出自己的观点,而针对不同立场进行论点的建构和陈述是很常用的议论文写作手法。收集和利用现有的针对某特定议题的讨论文本对于论文的写作至关重要。

随着不同领域的对话式论辩性文本的积累,针对对话式论辩文本的自动化理解工具和算法有越来越大的实际需求。经典的对话式论辩分析任务以两段论辩性文本作为输入(原始论辩文本和回复论辩文本),在经过针对每一个论辩性文本的论点抽取之后,深入的分析还包括三个子任务,即,交互论点对抽取,交互策略识别以及论辩质量评估。交互论点对抽取模块旨在将来自两段论辩性文本的论点进行自动化的匹配,交互策略识别则进一步研究观点对的交互性质(赞成、反对、驳斥等),论辩质量评估(有说服力、欠说服力等)。除了文本理解型的任务之外,生成式任务也在近年引起学界的广泛关注,研究主题包括总结性论点生成和目标论点生成。本综述将覆盖对话式论辩文本理解和生成两部分的研究内容,并介绍智能论辩应用的研究进展。最后,本文总结了当前的研究进展并展望了未来的发展方向。

1 交互式论点对抽取

1.1 交互式论点对识别

在对话式论辩中,一个核心的任务在于自动识别出论辩双方(或多方)关于某个特定议题、论点、或

论据,产生逻辑语义交互关系的论点对,即交互式论点对识别。从一方面来看,该任务是理解对话式论辩文本结构和内容的基础,只有识别出双方存在交互关系的论点对,才能进一步划分整场论辩中双方观点的相互作用和演化过程,从而推进更多对话式论辩中的下游任务的建模理解,如辩论摘要和论点说服力质量评估等;而从另一方面来说,该任务也是对话式论辩中最具有实际应用前景的方向,不论是在线论坛亦或是内容平台,识别出双方的观点交互后,有助于快速理解用户之间或买家卖家之间的核心争论点,从而帮助人们迅速捕捉讨论中涉及的多方诉求,精确定位到人们的真实意图和根本分歧。

交互式论点对识别任务最早由 Ji 等人在 2019 年提出。所谓交互式论点对,便是指在对话式论辩的场景中(如辩论赛或在线辩论论坛等),参与的双方就某一共同话题所产生的逻辑或语义上存在相关的论点对。图 1 给出了 ChangeMyView 论坛中的两条讨论帖的示例,其中,帖 A 为该主题下的原始帖(Original Post),即“楼主”所发。而帖 B 为 A 的回复帖(Reply Post)。这两篇帖子都旨在讨论联合养老制度的合理性,其中帖 A 支持联合养老保险制度,共分为五个论点句 $A_1 \sim A_5$,并从三个角度阐述了发帖者所认为的合理性;而帖 B 反对该制度,也分为五个论点句 $B_1 \sim B_5$,并通过举例论证的方式论证了自己的核心观点“人们退休后的收入需求通常会随着年龄的增长而下降”。通过分析这两篇帖子的文本,可以发现 B_1 和 A_1 之间存在直接的反对关系,因而这是一对交互式论点对,相似地, B_2 和 A_5 也共同成为一对交互式论点对。

Ji 等人通过观察分析 Tan 等人所爬取的 CMV 论坛数据,发现在该论坛数据中,一些回复帖中存在对于原始帖(被回复的帖子)的直接引用(Quotation)。图 2 是联合养老保险制度合理性讨论下的示例,可以发现帖 B 对帖 A 中“你把钱放在信托里,信托把钱投资出去,然后它就会和其他所有人的钱混在一起。”这句话进行了引用,且引用后紧跟的一句话便是在回复该被引用句。在论坛数据里,这样的引用由特殊符号“>”所标识。基于该发现,Ji 等人认为回复帖中关于原始帖引用后的第一句话便是对于原始帖被引用句的交互论点,即被引用句(Quotation Argument)和该回复句(Reply Argument)构成交互式论点对。因此,Ji 等人通过将 CMV 的文本数据进行规则筛选,共获得了 13 046 组引用回复论点对,即交互式论点对,并以此为正样本,按照 1:4 的比例在同一回复帖中采集其他

ational Auto Encoder)来对给定论点进行离散化表示,意在从糅合的论点文本表示中分离出论辩主题、

辩手行为特点等不同层面的特征,在所提数据集上的实验显示该模型相比其他的文本编码及匹配的基准模型有着较大的性能提升。随着该任务的提出,越来越多的研究者意识到交互式论点任务在对话式论辩中的重要地位,并提出了相关改进方法来提升模型性能^[33]。Yuan 等人^[34]通过对该任务数据集的分析,发现对该任务的建模不止需要对文本相似度进行计算,还需要对讨论中包含的概念实体和推理过程进行建模。基于该发现,作者基于 CMV 数据集构建了一个包含 20 余万节点和 80 余万条边的论辩领域知识图谱,将这一外部知识库引入模型增强上下文理解能力,并提出了基于 Transformer 编码器对推理路径进行建模来增强交互式论点识别任务的方法。

Devlin 等人^[35]提出了大规模预训练模型 BERT,凭借其庞大的参数和先进的多头自注意力机制,一时间在几乎所有的自然语言处理相关的任务上都取得了最优性能,在该任务上的表现也超过了 Ji 等人;Wu 等人^[36]使用混合提升(Mixup)方法,

通过对训练样本中的引用回复论点对及其训练目标进行线性插值,生成了大量的新的人造样本,同时作者还增强了所提模型对于给定引用论点和候选回复论点对上下文的建模,基于上述数据增强以及额外信息建模的改进使得其在该任务上模型的性能得到进一步的提升。

1.2 论点对抽取

论点对抽取(Argument Pair Extraction, APE)是对话式论辩领域的一个新任务,目的是从两篇相关的文章中成对抽取那些具有互动关系的论点。图 3 是同行评议场景下一个论点对抽取的例子,来自 Review-Rebuttal 数据集,左图是论文的审稿意见(Review),右图则是作者的回复(Rebuttal)。两篇文章在句子级别上被划分为论点和非论点。有阴影的为论点,没有阴影的为非论点。Review 中的论点可以与 Rebuttal 中的论点形成论点对,表示它们在讨论同一问题。在这个例子中,两个论点对分别用浅色阴影和深色阴影标注出。

Review	Arg		Arg	Rebuttal
这项工作将卷积神经网络应用于RGB-D室内场景分割任务	Non-Arg		Non-Arg	感谢您的审稿意见,这对我帮助很大。
该模型只是将深度作为单独的通道添加到卷积网络中的现有RGB通道中。			Rep: Arg-1	深度采集中的缺失值使用可用的修复代码进行预处理。
深度具有一些独特的属性,例如无穷大/缺失值,具体取决于传感器。	Rev: Arg-1	Arg-Pair-1		我们添加了对相关方法的论文引用。
可以添加一些关于如何正确整合深度的实验。			Rep: Arg-2	在论文中,我们假设深度未能超过RGB模型的类是深度图变化不大的对象类。
实验表明,利用深度信息的卷积网具有竞争力	Rev: Arg-2	Arg-Pair-2		现在,我们通过在图 2 中添加一些深度图来更好地强调这一观察结果。
...				
这是否表明深度并不总是有用的,或者可能有更好的方法来...				

图 3 同行评议场景下论点对抽取的示例^[37]

Cheng 等人^[38]指出,论点对抽取是一项非常有挑战的任务,其挑战性主要体现在以下两个方面:第一,从数据的层面看,不同于常见的抽取任务,该任务所面对的文本非常长,并且是两篇文章;第二,从任务定义的层面看,不同于传统的论辩关系预测任务,论辩对抽取首先需要从文本中抽取各个论点,然后再判断论点间的关系。

Bao 等人^[37]认为之前在 APE 任务中的方法通过两个分解的任务隐式地抽取出论点对,缺乏论点对之间参数级交互的显式建模,因而作者通过一个相互引导的框架来处理该论点对抽取任务,该框架可以利用一篇文章中的一个论点的信息来指导识别另一篇文章中与该论点成对的论点,以此方式使得

两篇文章可以在任务过程中相互引导,此外作者还引入了一个句子间关系图来描述评审和反驳文本句子之间的复杂交互作用,从而显式地利用论点对语义信息更精确地提取论点对。

Cheng 等人^[39]则认为以往 APE 任务上的研究将其看作是对直接连接在一起的两篇文章的序列标注问题和二分类问题,其局限性是没有充分利用两篇不同文章的独特特征和内在关系,从而提出了一种新的注意引导多层多交叉编码方案来解决这一问题,所提新模型处理两个独立序列编码器的两个段落,并通过注意相互更新它们的表示。此外,作者还通过更新两个序列的笛卡尔积表示,将预测部分表述为一个表填充问题,并引入了一

个辅助注意力损失,以引导每个论点对齐成对的论点。

2 论辩立场分析

立场是指一个人对某个目标(如物体、陈述或问题)的总体观点。对一个目标采取一种立场意味着支持或反对它。论辩立场分析是对文本给定目标的立场的分类。通常输入是一种论辩性文本,是问题或主张的某个目标,输出是这段文本的观点是赞成还是反对,有时也会考虑诸如中立或不相关的类型。论辩立场分析这个任务具有挑战性的原因在于:一是立场可以不提及话题本身就被表达;二是需要考虑所讨论的概念的对比性;三是积极的立场可以用消极的情绪来表达,反之亦然。但是立场和情绪的极性又往往是相关的。

对立场分类进行计算建模有两种方式:一是标准文本分类,用针对特定话题的文本进行训练;二是类似关系分类,以话题作为一个输入。常用文本特征如下:词袋,单词或单词 n -grams 的分布;核心词汇,主体性词汇的术语;POS,词性标签的分布;话语,论元间的连接词关系;情绪,以主题导向的情绪极性^[40-41]。

大多数立场分析方法是利用各种上下文信息,如相关论辩、讨论帖的文本等,使用有监督机器学习算法来学习主题,具体的立场分类方法如下:利用对话中作者的知识^[42]、利用对话中的对立观点^[41]、利用文本和话题目标的情感和对比^[43]、基于总体结构的立场识别路由核^[44]。

2010 年,Somasundaran 和 Wiebe^[40]探讨了情感和争论意见在意识形态辩论中分类立场的效用,建立了以情绪和争论观点及其目标为特征的有监督系统。2013 年,Hasan 和 Ng^[41]在辩论论坛上论辩双方的辩论文本的立场标签上建立用户交互约束和意识形态约束,从而检测辩手所表达的立场。2016 年,Sobhani 等人^[45]开发了一个立场检测系统,分析了包括情绪特征在内的各种特征对于立场检测的影响,并且在竞赛(SemEval-2016 task #6)中取得了优异的结果。2017 年,Bar-Haim 等人^[43]提出从情感和对比中得出立场,基于与给定话题相关的主张,对该主张在该问题上的立场进行分类。采用的方法过程如下:第一步,确定话题和主张的目标;第二步,针对每个目标的情绪极性进行分类;第三步,确定目标是否具有对比性;第四步,从情感和对比中得出立场。

前文介绍的立场检测,可以简单理解为是将文本中所表达的对给定目标的态度分类为“积极”“消极”或“中性”。这类立场检测工作大多是假设文中提到给定的目标,或者给出每个目标的训练数据。接下来介绍一些目标或主题未知情况下的立场分析工作。

2016 年,Augenstein 等人^[46]提出考虑立场检测任务中目标未知的情况,方法是:首先,需要学习一个模型,来解释文本对一个在文本中可能没有提到的目标的立场;其次,需要学习这样一个没有标记的目标的训练数据的模型。例如,一条推特“特朗普是共和党唯一诚实的声音”表达了对目标唐纳德·特朗普的积极立场。然而,当以希拉里·克林顿为隐含目标时,这条推特表达了一种负面的立场,因为支持一个政党的候选人意味着对其他政党的候选人的负面立场。在这个例子中,我们需要通过只使用其他目标的训练数据来学习希拉里·克林顿的模型。虽然这使得任务更具挑战性,但这是一个更现实的场景,因为不太可能有针对每个感兴趣的目标的标记训练数据。

2017 年,Wachsmuth 等人^[44]提出基于总体结构的立场识别核方法,给定一个单一的论辩性文本,在不知道所讨论的主题的情况下,对立场和偏见进行分类。假设论辩性文本的整体结构是决定立场和偏见的决定性因素,采用的方法是:从文本的论证结构开始,建模整体结构的路由核(树核的一个变体,可以捕获顺序结构和层次结构^[47]),再根据整体结构对立场和偏见进行分类。2020 年,Kobbe 等人^[48]提出了一个独立于主题的方法,其主要关注事件后果,即我们在表达观点时,经常提及某个事件发生的后果来支持我们的观点或反对他人的观点。该篇工作提出一种模式来匹配/建模这种表达方式,并直接根据模式的结果推理出其所表达的立场,实现无监督判断一个句子对于目标的立场。

当有足够的带标签的训练数据时,立场检测任务很容易取得成功。然而,注释足够的数据会耗费大量人力,这为将立场分类器泛化到具有新目标的数据设置了显著的障碍,且立场检测有不同的任务变体,任务定义也各不相同,包括标签类型(如有些标签既包括同意、不同意和不相关的变体,又包括难以交叉映射的变体,如讨论和问题)、数据集和标注规范不同。这些激发了跨目标立场检测的研究^[49],它通过利用来自源目标的大量注释数据来推断最终目标的态度。跨目标立场检测的方法一般是利用不同目标共享的常用词汇或概念级知识来弥合不同目标之间的知识

差距。2020 年,Zhang 等人^[50]提出了一种用于语义-情感知识转移(SEKT)模型用于跨目标立场检测,解决了基于概念的方法可能无法区分隐式带有术语和上下文信息的立场的问题。2021 年,Hardalov 等人^[51]设计了一个针对端到端的跨领域的标签自适应立场检测框架,模型基于来自预训练语言模型的输入表示,并使用混合专家和领域对抗训练来适应,用于预测看不见的、用户定义的域外标签。

3 论辩质量评估

3.1 论辩质量和论辩质量评估

论辩质量反映了一个单元、一个论点或论证有多好。例如,前提是否可接受、语言上是否清楚、文字是否与讨论有关、论证是否有说服力、说服是否有效、或论辩是否合理等。论辩质量评估需要考虑以下几点:一是目标导向性,哪个方面重要取决于论证的目标;二是颗粒度,质量评估可以在不同的文本颗粒度上进行处理;三是维度,评估时可能要综合多个质量维度。

论辩质量维度的分类有三个主要的方面,分别是逻辑性、修辞性、辩证性。逻辑性,指的是一个有说服力的论点要具有可接受的、相关的和充分的前提:局部可接受性,给出的前提值得被相信是真实的;局部相关性,该前提与结论相关;局部充分性,有这个前提就足以得出结论了。修辞性,指的是有效的论证,能够说服目标受众,包括如下几点:可信度,使作者值得被信任;情感吸引力,让听众愿意被说服;清晰度,语言上清晰,尽可能简单;适当性,语言上与听众和话题匹配;顺序性,以正确的行文顺序呈现内容。辩证性,指的是合理的论证,包括可接受的、相关的和充分的:全局可接受性,值得以陈述的方式加以考虑;全局相关性,有助于解决给出的话题或问题;全局充分性,充分反驳了潜在的反面意见^[6]。

论辩质量评估是识别论证中无可争辩的缺陷或要求,是对论证中某些质量维度的判断,确定论证是否成功地达到了其目标。例如,语言上是否清楚,说服是否有效。通常来说,论辩质量评估具有可选择性和主观性,即可以绝对地或相对地进行质量评估,且感知的质量取决于读者(听众)或作者(发言者)的看法。论辩质量评估的输入是有论证性的文本、元数据、外部知识等,现有的评估方法有有监督的分类/回归、基于图模型的分析等。

理论上,论辩质量评估要在一致性、合理性或相似性方面给出对论辩质量的标准观点,建议使用绝对质量评级。而在实践中,论辩质量是由对某些群体的有效性决定的,这意味着通常相对质量评级更合适。正所谓“一个有力的论据是一个有效的论据,可以赢得听众的支持,还是一个有效论据,应该赢得观众的支持?”^[52]

在某种意义上,关于论辩质量的问题是论辩挖掘的终极问题^[53]。这说明了论辩质量评估的重要性。那么为什么要评估论辩的质量呢?第一,在实践中我们对论辩挖掘和对论断的理解还不够充分;第二,对于成功的论辩,我们需要找出“最佳”的论点;第三,论辩质量评估对任何计算论辩的应用都至关重要。常见的应用有:论点搜索,哪个论点排名最高;写作支持,一个论辩文本有多好,是否存在什么缺陷;自动决策,哪些论点比其他论点更重要。

3.2 绝对质量评估和相对质量评估

绝对质量评估是从一个预定义的量表中分配一个分数作为评估结果。相对质量评估是给定两个实例,比较哪一个质量更好。相对质量评估通常要更容易,但绝对质量评估传播广泛,而且通常效果很好。

绝对质量评估要解决的问题是预测一个论点是否好(或有说服力、有效等)并给它的好坏评分。这里可以将质量评估视为一个标准的分类或回归任务,主要是学习哪项特征或元数据代表了论辩的质量。现有的方法有:支持力度,基于支持结论的证据数量^[54];说服力,基于参与者的交互作用来预测^[55];组织性和力度,基于结构的评估^[56];充分性,使用卷积神经网络进行分类^[57],基于生成的结论进行分类^[58]。

2016 年,Tan 等人^[55]提出基于交互作用的质量评估,分析语言特征和交互特征与说服的相关性,并根据说服是否会发生的特征进行预测,以研究在讨论中究竟是什么能说服那些愿意被说服的人。2016 年,Wachsmuth 等人^[56]提出基于论辩挖掘的质量评级,给出一篇有说服力的文章,给予论证相关的质量维度评分,以研究能否利用论辩挖掘来评估有说服力的论文的论证质量。该工作采用的质量维度包括:组织性,论证顺序安排得多有条理;论证的清晰度,即论文有多容易理解;一致性,文章与讨论的问题关系有多紧密;论证强度,支持这篇论文的论点有多有力^[59-62]。

然而,独立地评价一个论点的质量可能是困难的,甚至是不够可信的。相对质量评估就是一种更

简单或更现实来评估质量的方法,因为通常我们只对可用的最佳论点感兴趣,那么仅仅比较一个论点与其他论点的质量就足够了。目前的挑战是还无法确定选出的最好的论点是否足够好。现有的方法有:获胜方,从辩论流中预测辩论获胜方^[63],从内容和风格中预测获胜方^[64];说服力,SVM 和 Bi-LSTM 的辩论质量评估比较^[65];支持力度,根据支持主张对论点进行排名^[66]。

2016 年,Zhang 等人^[63]提出基于辩论流的质量比较,通过挖掘正反方的支持点,建模“会话流”(即一方什么时候提出自己的论点,什么时候攻击对方的论点),并用基于会话流特征的逻辑回归分类器,预测一场完整的牛津式辩论的哪一方会赢。2016 年,Habernal 和 Gurevych^[65]提出基于 SVM 和 Bi-LSTM 的辩论质量比较,用各种语言特征的非线性核支持向量机和双向长短时记忆神经网络,对给定的两个具有相同主题和立场的论点,预测哪一个更有说服力。

2017 年,Wachsmuth 等人^[6]提出代表理论的绝对质量评估数据,包含绝对专家评级、规范指南、15 个预定义的质量维度。2016 年,Habernal 和 Gurevych^[65]提出代表实践的相对质量评估数据,包含相对层面比较和 17+1 个结果标签,没有指南。2017 年,Wachsmuth 等人^[6]用 736 个带有评级和标签的论点对计算所有维度和结果的 Kendall 相关系数,对理论与实践的质量评估进行实证比较,结果表明,理论与实践的匹配程度大于预期;理论可以在实践中指导质量评估;实践指出了要注意哪些重点来简化理论。

3.3 客观质量评估和主观质量评估

质量评估具有主观性。首先,在许多维度上质量评估本质上都是主观的;其次,质量取决于一个问题的不同方面的主观权重。同时,质量评估也依赖于先入之见。如关于死刑的两个论点,“死刑使一种不可逆转的暴力行为合法化。只要人类的判断仍然容易出错,处决无辜者的风险就永远无法消除。”“死刑并不能阻止人们犯下严重的暴力犯罪。只有被抓住并受到惩罚才令人沮丧。”哪个与主题更相关呢?有两种方法可以解决这个问题,一是关注可以被评估为“客观”的属性,二是在质量评估中包括一个读者或听众的模型。

客观质量评估要解决的问题是如何在不学习主观注释的情况下评估质量,以及什么是客观的质量

指标。其主要思想是基于所有论点所产生的结构来评估质量,适用于绝对评估和相对评估。其面临的一大挑战是对主观注释的评价的处理,可能的解决方案是依赖于对许多注释者的多数评估。现有的方法有:可接受性,基于攻击关系的评估^[67];相关性,基于单元重复利用的评估^[68];突出性,基于论点频率的评估^[69]。

2012 年,Cabrio 和 Villata^[67]提出基于攻击关系的客观评估,给定一组论点,对攻击进行识别,并根据 Dung 提出的框架^[70]评估论点的可接受性。Dung 于 1995 年提出的抽象论证框架是一个有向图,其中节点表示论点,边表示论点之间的攻击关系,揭示了是否接受一个论点。2017 年,Wachsmuth 等人^[68]提出基于单元重复利用的客观评估,研究给定一组论点,哪一个与某些话题最相关,然而相关性是高度主观的,即寻找一个“客观”的相关性度量。该工作假设一个结论的相关性取决于网络上的其他论点将其作为前提,暂时忽略论点的内容和推理,从网络规模上的结论重用中获得结构上的相关性。

主观质量评估要考虑的问题是,有效的论证最终都需要考虑目标受众。如果不这样做,人类几乎不需要辩论。其主要思想是在质量评估过程中建模目标受众,包括特定于受众的正确标注。然而,到目前为止,受众模型很少被明确地包括在研究方法内,且一些带标注的语料库实际上可能代表特定的受众。现有的方法有:不同的个性,感性论证与理性论证的有效性^[71];不同的意识形态,新闻社论的有效性^[72]。

2017 年,Lukin 等人^[71]提出基于个性的有效性评估,假设不同个性的人愿意接受不同类型的论点,研究五大个性(开放性、自觉性、外向性、一致性和精神性)对感性论证与理性论证有效性的影响。2018 年,El Baff 等人^[72]提出基于意识形态(分为保守派和自由派)的有效性评估,假设先验立场取决于政治意识形态(和人格),研究意识形态(和人格)对新闻社论有效性的影响,即是否挑战或加强立场。

4 对话式论辩生成

对话式论辩另一个分支的研究则偏向于提出自动化模型在对话式领域实现对话生成任务。对话生成任务是人机交互中机器实现输出的一个重要组成部分,如何让机器按照我们预先设定的范式实现有效输出是重中之重。具体的,对话生成任务大致可以分为总结性论点生成和目标论点生成两类。

4.1 总结性论点生成

最简单的对话式论辩生成任务之一是根据给定的观点或论点文本集合,生成对应的总结性句子,相当于生成了一个特定的论点句。从别人那里收集意见是我们日常活动的一个组成部分。发现别人的想法可以帮助我们在生活的不同方面导航,从日常任务的决定到判断基本的社会问题和形成个人意识形态。为了有效地吸收大量固执己见的信息,迫切需要自动化系统对一个实体或话题生成简洁流畅的意见总结。尽管在意见总结方面有大量的研究,但最突出的方法主要是采用抽取式摘要方法,即从原始文献中选择短语或句子纳入摘要^[73]。

Wang 等人^[74]从烂番茄网站上爬取影评并据此构建了一个影评数据集,其中包括了 3 731 部电影和 246 164 条评论,同时每部电影都额外包含一句评价作为基准。作者研究了为固执己见的文本生成摘要的问题。本文提出了一种基于注意力的神经网络模型,能够从多个文本单元中吸收信息,构建信息丰富、简洁、流畅的摘要。一种基于重要性的采样方法被设计用来允许编码器集成来自输入的一个重要子集的信息。自动评估表明,作者设计的系统在两个新收集的电影评论和论点数据集上的性能优于最先进的抽象和提取摘要系统。而作者的系统摘要在人类评价中也被评为信息量更大、语法更规范的。

Li 等人^[75]重点研究了综述总结的方法。不同于以往大多数采用语言规则或统计方法的研究,作者将审查挖掘任务定义为一个联合结构标注问题,提出了一种基于条件随机场(CRFs)的机器学习框架。它可以利用丰富的特征联合提取复习句的积极观点、消极观点和客体特征。语言结构可以自然地融入到模型表示中。除了线性链结构,作者还研究了连接结构和句法树结构。对电影评论和产品评论数据集的广泛实验表明,结构感知模型优于许多当时最先进的论辩挖掘方法。

Syed 等人^[76]为了进一步实现长文本总结,构建了一个“论辩文本+中心论点”的大型语料库 WebisConcluGen-21。作者研究了两种生成结论的范式;一种是提炼的,另一种是抽象的。后者利用论证知识,通过控制代码来增加数据,并在语料库的几个子集上微调 BART 模型。作者深入分析了语料库对任务的适用性、两代范式之间的差异、信息性和简便性之间的权衡以及编码论证知识的影响。其语料库、代码和训练的模型都是公开的。

4.2 目标论点生成

针对特定的话题生成支持或反对的高质量论点文本,则是更为实际的论点生成任务。针对生成任务的早期工作,通常是先从语料库中提取文本,然后按特定顺序输出,缺少综合不同语料内容的能力。产生高质量的论点在决策和推理过程中起着至关重要的作用。许多最终决策都是在争论或反驳中不断推进发展的,而当这种争论来到人机交互场景时,生成高质量论点的重要性就体现出来了。例如,立法机构经常进行辩论,以确保法案获得足够的票数通过;网上审议则是另一个常见的场景,它已经成为征求公众意见的一种流行方式。尽管如此,构建有说服力的论点对人类和计算机来说都是一项艰巨的任务。

Hua 等人^[77]首先提出了目标论点生成这一任务,任务被构建为建立这样一个模型:能够针对给定的论述,自动化生成不同立场的论点。作者提出了一个基于神经网络和编码-解码器结构的论点生成模型,丰富了从维基百科外部检索的论据。在 Reddit 上收集的大规模数据集上的实验表明,根据自动评估和人工评估,提出的模型比流行的序列到序列生成模型构建了更多与主题相关的内容。

Alshomary 等人^[78]重点关注了反论点的生成任务。此前的方法主要集中在反驳一个给定的结论,而他们研究的场景并不局限于此。考虑到识别论点的薄弱前提是有有效反击的关键,作者探索了破坏论证任务,也就是通过攻击一个论证的前提来反驳一个论证。下图是破坏论证的一个基本逻辑结构,机器首先识别论证结构中的弱前提,再针对弱前提生成反论点。具体的,作者提出了一种管道方法,首先评估这些前提的强度,然后针对较弱的前提提出反论点。一

论点: 女权主义正处于第三波浪潮中, 它在美国等国家造成了大量的麻烦。

前提: 我将提出我自己听到的几个女权主义论点。首先, 男女之间的工资差距很可怕。事实上, 自1960年代初以来, 向妇女支付低于男子的工资是非法的。其次是粉红税。女性产品的价格当然会高于男性。他们使用完全不同的化学物质来迎合更柔软的皮肤.....第四点, 我们承认女性运动有些时候是有效且有益的, 特别是在伊拉克、印度和沙特阿拉伯这些女人被迫去遮盖整个身体的国家。但是女性运动却首先发源于那些不会因为嫁妆卖女儿, 也没有受到类似压迫的国家中, 这是彻头彻尾的荒谬。

反论点: 事实上, 某些国家女性的情况更糟并不意味着女性在另一些国家的处境一定很好。例如美国50个州中有29个州会以同行恋行为解雇职员。而在文莱, 人们因同性恋而被石头砸死的事实, 更是同性恋恐惧症的一个例子.....

图4 破坏论证的基本逻辑结构

方面,人工评估和自动评估都证明了识别弱前提在反论证生成中的重要性。另一方面,在考虑正确性和内容丰富性时,相比传统方法,人工仲裁更喜欢这种基于弱前提检索的反论点生成方法。

Hua 等人^[79]为了解决传统论点生成模型中经常产生与输入不连贯和不忠实的输出的问题,提出了一个端对端训练的两步生成模型,该模型能够实现句级别的内容规划和风格规范。其中句子级别的内容规划首先决定要涵盖的关键短语以及所需的语言风格,然后是一个表面实现解码器,生成相关和连贯的文本。模型主要考虑了三个不同维度上的子任务——Reddit 中有说服力的论点构建,维基百科的简介生成以及科学文章的摘要生成。自动评估表明,作者提出的系统可以大大超过竞争比较。不仅如此,与不考虑语言风格的变体相比,人工仲裁认为该系统生成的文本更合理,表达更流畅。

有时我们不仅要限制目标论点的主题,还会限制论点讨论的具体角度。Schiller 等人^[80]提出了一个论点生成模型 Arg-CTRL。这一模型可以为给定的主题、立场和角度生成句子级别的论点。作者定义了论证角度检测问题,并认为这一问题是实现针对论证角度的细粒度控制的必要方法,并将 5 032 个带论证角度注释的辩论文本集合为一个数据集。实验表明,Arg-CTRL 模型能够生成高质量的、从特定角度切入的论点,尤其适用于自动生成反论点。

考虑到知识图在支持一般文本生成任务中的有效性,Al-Khatib 等人^[81]研究了论证相关知识图在控制论证生成中的应用。在该研究中,作者构建并填充了三个知识图,利用它们的几个组成部分将各种知识编码到辩论门户的文本和维基百科的相关段落中。具体操作上,作者使用编码知识的文本来微调预先训练的文本生成模型 GPT-2。模型有效性实验涵盖了论证环境中的几个重要维度,包括论证性和似是而非性,手动和自动地评估新创建的论证。结果表明,从编码图表的知识到辩论门户文本产生的积极影响比那些没有知识产生的论点质量更好。

Wachsmuth 等人^[82]探究了语法修辞策略。好的论证不能仅有严密的逻辑结构,还需要有合理的语言结构以及有温度的表达。作者研究了在论点生成时,如何将逻辑推理与对伦理和情感的诉求结合起来。这一策略意味着如何选择和安排机器抽取的论点并将它们有效地串联起来。在建模时,作者让 26 位专家针对 10 个主题,用不同的策略合成议论文本。研究发现,当采用相同的策略时,专家对选择

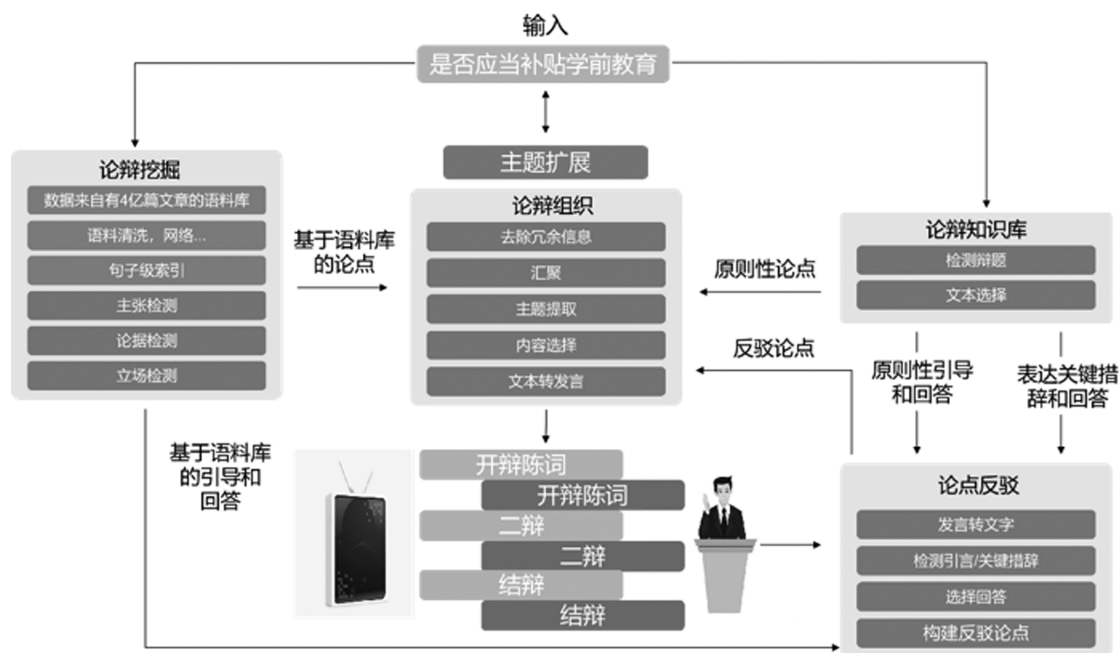
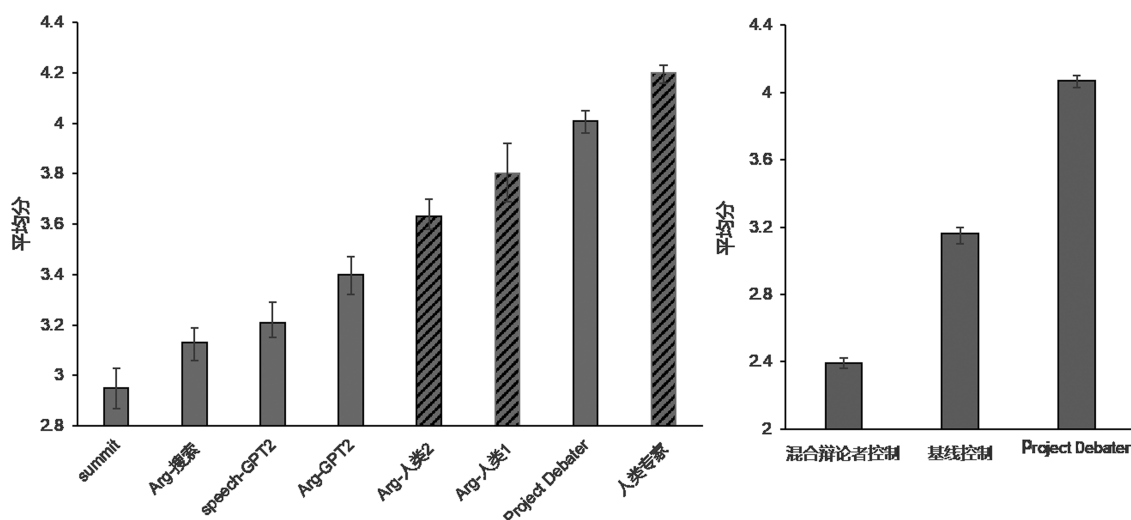
的同意程度明显更高。这意味着作者采用的语法修辞策略是成功的,因为虽然不同文章的文本有显著差异,但它们的安排保持稳定,也即遵循了同一套语法修辞策略。

5 智慧论辩应用

IBM 于 2019 年公开发布了人工智能辩手 Project Debater^[83]。Project Debater 是全世界首个能与人类进行复杂辩论的自动化论辩系统。该项目由 IBM 团队自 2012 年启动开发,2021 年 3 月登上了《Nature》杂志的封面。2019 年 2 月 11 日 Project Debater 与 H. Natarajan(纳塔拉扬,世界大学生辩论赛冠军)围绕“是否应当补贴学前教育”展开了一场公开辩论,人类辩手持正方,AI 辩手持反方。比赛采用简化后的议会制辩论,含 15 分钟持题准备时间,三轮交替发言环节。赛前,79%的听众同意学前教育应该得到补贴,13%的人不同意。赛后,62%的人同意,30%的人不同意。最终,人类辩手 H. Natarajan 获得胜利。

如图 5 所示,Project Debater 系统包含论辩挖掘、论辩知识库、论点反驳和论辩组织四个模块^[84]。论辩挖掘模块从大的文本语料库中找寻到议题相关的论点和驳论点。论辩知识库包含论点、驳论点以及其他辩题下的相关文本;一旦给定辩题,系统在其中找到最相关的论辩语料。论点反驳模块将前两个模块中潜在的相反论点与实际对手的陈词做匹配,由此生成可能的回应。最后,论辩组织模块从其他模块提供的文本中选择性地组织出一则连续的发言。

图 6 展示了 Project Debater 的性能评估结果。左图展示了 Project Debater 与其他基线系统的对比。条形表示平均分数,其中 5 表示对“这篇演讲是支持该主题的良好开场演讲”的观点“非常同意”,1 表示“非常不同意”。带斜线的条形图表示该系统中的语音是由人类生成的或依赖于人工编写的论点。右图展示了最终系统的评估结果。“Project Debater”描述了 Project Debater 生成 S1 和 S3 时的结果。在“混合辩论者控制”中,第三次演讲是由 Project Debater 在另一个辩题生成下的 S3。在“基线控制”中,S1 和 S3 都是从全自动基线系统之一中选择的开场白。条形表示平均分数,其中,5 表示对于“第一个发言者在这场辩论中表现得不错”的观点“非常同意”,1 表示“非常不同意”。Project Debater 的结果明显优于其他所有基准,并且非常接近人类专家的分數。

图 5 IBM Project Debate 框架图^[83]图 6 IBM Project Debater 评测结果^[83]

为支撑该系统的搭建,项目团队在论辩挖掘、语音理解与生成、文本生成等多项子任务方面进行了探索研究,构建了大量优质论辩子任务数据集,研究成果公开发表在 ACL、EMNLP 等权威会议中。项目公开了用于论辩系统构建的 API 以及大量子任务数据集,包括主张检测、主张边界检测、证据检测、论点质量评估、立场识别、关键点评估及立场生成等,广泛用于学界研究^[84]。

6 总结和展望

用语言和逻辑的角度分析辩论,一直是人们探

寻辩论背后人类智慧规律的重要方法;而计算机、机器学习、人工智能等新技术的不断发展,无疑让计算论辩走上了发展的快车道,也让越来越多的计算论辩成果落地成为可能。

面临的挑战 作为一个源远流长但直到最近才以一个整体为人们所关注的研究领域,目前计算论辩仍然存在一些挑战,有待后续研究攻坚克难。

(1) **缺乏通用的大型标准数据集** 近年来,不断有新的计算论辩任务涌现出来,扩充着这一子领域的谱系。这些新兴任务的相关工作通常都缺乏既有的数据集,因此不得不自行标注并构建用于训练、测试的小型数据集。大规模标准评测数据集的空

缺,使有关研究提出的模型无法在大型语料上验证效果,同时又导致更多的小数据集出现,却难以将它们统合为一个大型数据集。

(2) **尚未形成一套完整的研究范式** 计算论辩的研究扎根于论辩分析理论,然而正如前文所述,目前与论辩相关的背景理论繁多,相互之间各有所长,却并没有形成能覆盖绝大多数论辩场景的统一理论。采用不同理论基础的研究工作往往会发展出不同的研究范式,这就为特定方向上各个研究的横向对比与融合增添了阻碍。

发展趋势 当下计算论辩也展现出一些有趣的发展趋势,其中的一个或多个有可能成为未来这一领域的研究主流。

(1) **基准评测数据出现**,为计算论辩提供数据基石。虽然我们还没能构建出一个普适的大型基准数据集,但如今人们每天都在生产大量论辩语料。在许多研究人员和标注人员的不懈努力下,它们也在不断衍生出各类数据量大、任务齐全、语种齐全的计算论辩专用数据。这些工作使得将来大型基准评测数据集的构建成为可能。

(2) **小样本学习、领域迁移方法成为研究热点**。作为自然语言处理的一个分支,在各种 NLP 任务中受到关注的小样本学习和领域迁移方法自然也不会缺席计算论辩相关研究。事实上,许多特殊形式的论辩语料(如庭审记录)并不容易大量获取,而且在相似的论辩框架下可以蕴含千万种语义信息,因此如何利用有限且有局限的论辩数据学习背后的论辩框架,是目前值得研究的热点课题。

(3) **在逻辑判定之外,价值属性开始凸显**。过去的计算论辩研究多数注重论点本身蕴含的语义和逻辑,对论点背后辩方的价值取向关注不多。但在社交媒体高度发达的今天,社交平台上的许多交锋实质上体现了不同群体之间的价值观与意识形态冲突。因此,越来越多的研究开始探讨论辩文本中蕴含的价值属性,即所谓的价值观辩论或意识形态建模。

(4) **多模态信息的相关研究正在引起人们的关注**。大多数的辩论信息都以论辩文本或语料的形式呈现,但在许多情景(特别是线下的面对面辩论)中,辩论双方的声学特征、面部表情、肢体语言等非文本信息实际上都蕴含一定的信息量,并且会影响辩论的质量与结果。基于上述观察,最近的一些研究开始建立利用多模态信息的计算论辩技术,并在论辩质量评估等任务上取得了一定成果。

此外,还有许多新的方向与课题,例如,论辩信

息的图谱表示、群体语境下的自主论辩等,它们都有机会在接下来的数年时间里发展为计算论辩的又一个闪光点。无论如何,在如今这个充满着观点对立与信息茧房的社交媒体时代,已然发展出多种新的辩论形式,而计算论辩在这个时代迸发的无穷潜力,依然等待着人们的努力挖掘。

参考文献

- [1] VAN EEMEREN FH, JACKSON S, JACOBS S. Argumentation[M]//Reasonableness and Effectiveness in Argumentative Discourse. Springer, Cham, 2015: 3-25.
- [2] FREELEY AJ, STEINBERG DL. Argumentation and debate[M]. Cengage Learning, 2013.
- [3] BESNARD P, HUNTER A. Elements of argumentation[M]. Cambridge: MIT Press, 2008.
- [4] VAN EEMEREN FH, Grootendorst R, Henkemans AF. Argumentation: Analysis, evaluation, presentation[M]. Routledge, 2002.
- [5] EGER S, DAXENBERGER J, GUREVYCH I. Neural end-to-end learning for computational argumentation mining[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017: 11-22.
- [6] WACHSMUTH H, NADERI N, HOU Y, et al. Computational argumentation quality assessment in natural language[C]//Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics, 2017: 176-187.
- [7] RAHWAN I, SIMARI GR. Argumentation in artificial intelligence[M]. Heidelberg: Springer, 2009.
- [8] BENCHCAPON TJ, DUNNE PE. Argumentation in artificial intelligence[J]. Artificial Intelligence, 2007, 171(10-15): 619-641.
- [9] ALKHATIB K, WACHSMUTH H, HAGEN M, et al. Crossdomain mining of argumentative text through distant supervision[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016: 1395-1404.
- [10] PELDSZUS A, STEDEM. Joint prediction in mst-style discourse parsing for argumentation mining[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2015: 938-948.
- [11] STAB C, GUREVYCH I. Identifying argumentative discourse structures in persuasive essays[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2014: 46-56.
- [12] FENG V W, LIN Z, HIRST G. The impact of deep hier-archical discourse structures in the evaluation of text coherence[C]//Proceedings of COLING, the 25th International Conference on Computational Linguistics: Technical Papers. 2014: 940-949.
- [13] WYNER A Z. A functional perspective on argumen-

- tation schemes[J]. *Argument & Computation*, 2016, 7(2-3): 113-133.
- [14] TAGHIPOUR K, NG H T. A neural approach to automated essay scoring[C]//*Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2016: 1882-1891.
- [15] MIM F S, INOUE N, REISERT P, et al. Unsupervised learning of discourse-aware text representation for essay scoring[C]//*Proceedings of the 57th Conference of the Association for Computational Linguistics*, 2019: 378-385.
- [16] ASTERHAN CS, SCHWARZ BB. The effects of monological and dialogical argumentation on concept learning in evolutionary theory[J]. *Journal of Educational Psychology*, 2007, 99(3): 626.
- [17] HUNTER A, POLBERG S. Empirical methods for modelling persuadees in dialogical argumentation. [C]//*Proceedings of the 29th International Conference on Tools with Artificial Intelligence*, 2017.
- [18] SHIRAFUJI D, RZEPKA R, ARAKI K. Debate outcome prediction using automatic persuasiveness evaluation and counterargument relations[C]//*Proceedings of the IJCAI Workshop on Linguistic and Cognitive Approaches To Dialog Agents Workshop LACATODA*, 2019.
- [19] IMOSSATEN A, MONTMAIN J, RICO A, et al. A dynamical model for simulating a debate outcome [C]//*Proceedings of the 3rd International Conference on Agents and Artificial Intelligence*, SciTe-Press, 2011: 31-40.
- [20] BESNARD P, GARCÍA A J, HUNTER A, et al. Introduction to structured argumentation[J]. *Argument & Computation*, 2014, 5(1): 1-4.
- [21] COCARASCU O, TONI F. Combining deep learning and argumentative reasoning for the analysis of social media textual content using small data sets[J]. *Computational Linguistics*, 2018, 44(4): 833-858.
- [22] SCHNEIDER J, GROZA T, PASSANT A. A review of argumentation for the social semantic web[J]. *Semantic Web*, 2013, 4(2): 159-218.
- [23] ALAGOZ E. Social argumentation in online synchronous communication [J]. *International Journal of Computer Supported Collaborative Learning*, 2013, 8(4): 399-426.
- [24] SARTOR G. Introduction: Judicial applications of artificial intelligence [J]. *Artificial Intelligence and Law*, 1998, 6(2-4): 105-110.
- [25] CARBONELL F. Reasoning by consequences: Applying different argumentation structures to the analysis of consequentialist reasoning in judicial decisions [M]. Springer Netherlands, 2013.
- [26] MAZZI D. The construction of argumentation in judicial texts: Combining a genre and a corpus perspective[J]. *Argumentation*, 2007, 21(1): 21-38.
- [27] ERDURAN S, JIMÉNEZ ALEIXANDRE MP. Argumentation in science education. Perspectives from classroom-Based Research[M]. Springer, 2008.
- [28] JIMÉNEZ ALEIXANDRE MP, ERDURAN S. Argumentation in science education: An overview[J]. *Argumentation in Science Education*, 2007, 35: 3-27.
- [29] DUSCHL R A, OSBORNE J. Supporting and promoting argumentation discourse in science education[J]. *Studies in Science Education*, 2002, 38(1): 39-72.
- [30] MIRZA NM, PERRETCLERMONT AN. Argumentation and education: Theoretical foundations and practices[M]. Springer Science & Business Media, 2009.
- [31] BURNSTEIN E, VINOKUR A. Persuasive argumentation and social comparison as determinants of attitude polarization[J]. *Journal of Experimental Social Psychology*, 1977, 13(4): 315-332.
- [32] JI L, WEI Z, LI J, et al. Discrete argument representation learning for interactive argument pair identification[C]//*Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language*, 2021: 5467-5478.
- [33] YUAN J, WEI Z, GAO Y, et al. Overview of SMP-CAIL2020 Argmine: The interactive argument pair extraction in judgement document challenge[J]. *Data Intelligence*, 2021: 1-20.
- [34] YUAN J, WEI Z, ZHAO D, et al. Leveraging argumentation knowledge graph for interactive argument pair identification[G]//*Proceedings of the Association for Computational Linguistics: ACL-IJCNLP*, 2021: 2310-2319.
- [35] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//*Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Minneapolis, Minnesota: Association for Computational Linguistics, 2019: 4171-4186.
- [36] WU Y, LIU P. ACE: A context enhanced model for interactive argument pair identification[C]//*Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing*, 2021: 569-578.
- [37] BAO J, LIANG B, SUN J, et al. Argument pair extraction with mutual guidance and intersentence relation graph [C]//*Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2021: 3923-3934.
- [38] CHENG L, BING L, YU Q, et al. APE: Argument pair extraction from peer review and rebuttal via multi-task learning[C]//*Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2020: 7000-7011.
- [39] CHENG L, WU T, BING L, et al. Argument pair extraction via attention-guided multilayer multi-cross encoding[C]//*Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 2021: 6341-6353.
- [40] SOMASUNDARAN S, WIEBE J. Recognizing stances in ideological online debates[C]//*Proceedings of the NAACL HLT Workshop on Computational*

- Approaches to Analysis and Generation of Emotion in Text, 2010: 116-124.
- [41] HASAN K S, NG V. Stance classification of ideological debates: Data, models, features, and constraints [C]//Proceedings of the 6th International Joint Conference on Natural Language Processing, 2013: 1348-1356.
- [42] RANADE S, SANGAL R, MAMIDI R. Stance classification in online debates by recognizing users' intentions [C]//Proceedings of the SIGDIAL Conference, 2013: 61-69.
- [43] BARHAIM R, BHATTACHARYA I, DINUZZO F, et al. Stance classification of context-dependent claims [C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2017: 251-261.
- [44] WACHSMUTH H, DA SAN MARTINO G, KIESEL D, et al. The impact of modeling overall argumentation with tree kernels [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2017: 2379-2389.
- [45] SOBHANI P, MOHAMMAD S, KIRITCHENKO S. Detecting stance in tweets and analyzing its interaction with sentiment [C]//Proceedings of the 5th Joint Conference on Lexical and Computational Semantics, 2016: 159-169.
- [46] AUGENSTEIN I, ROCKTÄSCHEL T, VLACHOS A, et al. Stance detection with bidirectional conditional encoding [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2016: 876-885.
- [47] AIOLLI F, DA SAN MARTINO G, SPERDUTI A. Route kernels for trees [C]//Proceedings of the 26th Annual International Conference on Machine Learning, 2009: 17-24.
- [48] KOBBE J, HULPU TEXTCOMMABELOWS I, STUCKENSCHMIDT H. Unsupervised stance detection for arguments from consequences [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2020: 50-60.
- [49] WEI P, MAO W. Modeling transferable topics for cross-target stance detection [C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019: 1173-1176.
- [50] ZHANG B, YANG M, LI X, et al. Enhancing cross-target stance detection with transferable semantic-emotion knowledge [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020: 3188-3197.
- [51] HARDALOV M, ARORA A, NAKOV P, et al. Cross domain label adaptive stance detection [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2021: 9011-9028.
- [52] PERELMAN C, OLBRECHTSTYTECA L. The New Rhetoric: A treatise on argumentation [M]. Notre Dame University Press, 1969.
- [53] CHERSONI E. Manfred stede and jodi schneider; Argumentation mining, synthesis lectures on human language technologies [J]. Lang Resour Eval, 2021, 55(3): 869-871.
- [54] RAHIMI Z, LITMAN D J, CORRENTI R, et al. Automatic scoring of an analytical response to text assessment [C]//Proceedings of the 12th International Conference on Intelligent Tutoring Systems, 2014: 601-610.
- [55] TAN C, NICULAE V, DANESCU NICULESCU MIZIL C, et al. Winning Arguments: Interaction dynamics and persuasion strategies in good-faith online discussions [C]//Proceedings of the 25th International Conference on World Wide Web, 2016: 613-624.
- [56] WACHSMUTH H, ALKHATIB K, STEIN B. Using argument mining to assess the argumentation quality of essays [C]//Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers, 2016: 1680-1691.
- [57] STAB C, GUREVYCH I. Recognizing insufficiently supported arguments in argumentative essays [C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2017: 980-990.
- [58] GURCKE T, ALSHOMARY M, WACHSMUTH H. Assessing the sufficiency of arguments through conclusion generation [C]//Proceedings of the 8th Workshop on Argument Mining, 2021: 67-77.
- [59] PERSING I, DAVIS A, NG V. Modeling organization in student essays [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2010: 229-239.
- [60] PERSING I, NG V. Modeling thesis clarity in student essays [C]//Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 2013: 260-269.
- [61] PERSING I, NG V. Modeling prompt adherence in student essays [C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014: 1534-1543.
- [62] PERSING I, NG V. Modeling argument strength in student essays [C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015: 543-552.
- [63] ZHANG J, KUMAR R, RAVI S, et al. Conversational flow in Oxford style debates [C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016: 136-141.
- [64] WANG L, BEAUCHAMP N, SHUGARS S, et al. Winning on the Merits: The joint effects of content and style on debate outcomes [G]//Transactions of the Association for Computational Linguistics, 2017, 5: 219-232.
- [65] HABERNAL I, GUREVYCH I. Which argument is more convincing? Analyzing and predicting convincingness of Web arguments using bidirectional LSTM [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 1589-1599.

- [66] BRAUNSTAIN L, KURLAND O, CARMEL D, et al. Supporting human answers for advice seeking questions in CQA sites[C]//Proceedings of the European Conference on Information Retrieval, 2016.
- [67] CABRIO E, VILLATA S. Combining textual entailment and argumentation theory for supporting online debates interactions [C]//Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 2012: 208-212.
- [68] WACHSMUTH H, STEIN B, AJJOUR Y. "PageRank" for argument relevance [C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2017: 1117-1127.
- [69] BOLTUŽIĆ F, ŠNAJDER J. Identifying Prominent Arguments in online debates using semantic textual similarity [C]//Proceedings of the 2nd Workshop on Argumentation Mining, 2015: 110-115.
- [70] DUNG P M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and N -preson games [J]. Artificial Intelligence, 1995, 77(2): 321-357.
- [71] LUKIN S, ANAND P, WALKER M, et al. Argument strength is in the eye of the beholder: Audience effects in persuasion [C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2017: 742-753.
- [72] EL BAFF R, WACHSMUTH H, ALKHATIB K, et al. Challenge or empower: Revisiting argumentation quality in a news editorial corpus [C]//Proceedings of the 22nd Conference on Computational Natural Language Learning, 2018: 454-464.
- [73] Hu M, Liu B. Mining opinion features in customer reviews[C]//Proceedings of the 19th National Conference on Artificial Intelligence, 2004: 755-760.
- [74] WANG L, LING W. neural network-based abstract generation for opinions and arguments [C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016: 47-57.
- [75] Li F, Han C, Huang M, et al. Structure-aware review mining and summarization[C]//Proceedings of the 23rd International Conference on Computational Linguistics. Association for Computational Linguistics, 2010: 653-661.
- [76] SYED S, ALKHATIB K, ALSHOMARY M, et al. Generating informative conclusions for argumentative texts[C]//Proceedings of the Association for Computational Linguistics, 2021: 3482-3493.
- [77] HUA X, WANG L. Neural argument generation augmented with externally retrieved evidence[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 219-230.
- [78] ALSHOMARY M, SYED S, POTTHAST M, et al. Argument undermining: Counter-Argument generation by attacking weak premises[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 2661-2672.
- [79] HUA X, WANG L. Sentence-level content planning and style specification for neural text generation[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019: 591-602.
- [80] SCHILLER B, DAXENBERGER J, GUREVYCH I. Aspect controlled neural argument generation[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021: 380-396.
- [81] AL KHATIB K, TRAUTNER L, WACHSMUTH H, et al. Employing argumentation knowledge graphs for neural argument generation[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021: 4744-4754.
- [82] WACHSMUTH H, STEDE M, BAFF R E, et al. Argumentation synthesis following rhetorical strategies[C]//Proceedings of the 27th International Conference on Computational Linguistics, 2018: 3753-3765.
- [83] IBM Research, Project Debater [CP/OL]. <https://research.ibm.com/interactive/project-debater/>. [2022-08-24].
- [84] SLONIM N, BILU Y, ALZATE C, et al. An autonomous debating system [J]. Nature, 2021, 591(7850): 379-384.



魏忠钰(1987—),通信作者,博士,副教授,主要研究领域为自然语言处理、社会计算、论辩挖掘。
E-mail: zywei@fudan.edu.cn



沈晨晨(1995—),硕士研究生,主要研究领域为论辩挖掘、领域大模型微调。
E-mail: ccshen22@m.fudan.edu.cn



丁佳玟(1998—),硕士研究生,主要研究领域为自然语言问答。
E-mail: jyding20@fudan.edu.cn