

论辩挖掘研究

宋巍¹, 魏忠钰²

¹首都师范大学, ²复旦大学

个人简介:

宋巍, 首都师范大学信息工程学院, 讲师, 中国中文信息学会青工委委员, 在哈尔滨工业大学计算机系获得学士、硕士和博士学位。研究方向为信息检索与自然语言处理, 主要研究兴趣包括用户分析、文本篇章分析与质量评估以及面向教育领域的自然语言处理等, 在 SIGIR, WWW, COLING, EMNLP 等一流与重要国际会议及相关期刊发表论文 10 余篇。

魏忠钰, 复旦大学大数据学院, 青年副研究员, 中国中文信息学会青工委委员, 美国德州大学达拉斯分校博士后, 博士毕业于香港中文大学, 在哈尔滨工业大学取得学士和硕士学位。从事自然语言处理, 社交媒体分析, 论辩挖掘等方面的研究, 在 SIGIR, ACL, COLING 等国际一流与重要会议发表论文 10 余篇。

在刚刚结束的国际计算语言学会议 COLING 2016 上, 剑桥大学的 Simone Teufel 教授以计算论辩(Computational Argumentation)为主题进行了大会报告, 获得强烈反响。近年来, 自然语言处理顶级会议 ACL 以及人工智能顶级会议 IJCAI 均曾开设关于论辩挖掘(Argumentation Mining)的讲习班或 Workshop。~~不经意间, 论辩一词开始频繁地出现在我们的视野里。~~本文将简要介绍论辩挖掘及相关工作。

1 引言

论辩(Argumentation)研究辩论和推理的过程, 是一个涉及到逻辑、哲学、语言、修辞、法律和计算机科学等多学科的研究领域。在人工智能领域研究论辩激发生了一个新的研究方向——计算论辩(Computational Argumentation) [1]。计算论辩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提高人工智能自动推理的能力。

论辩挖掘(Argumentation Mining)是计算论辩中一个重要的任务, 它的主要目标是自动地从文本中提取论点(Argument), 以便为论辩和推理引擎的计算模型提供结构化数据。这里涉及的几个术语很容易让人迷惑。我们倾向于将论点(Argument)视为一种结构, 用论辩(Argumentation)强调构建论点结构的过程, 将论辩挖掘(Argumentation Mining)视为从数据中发现结构化论点的任务。例如, 一种典型的论点结构包括前提(Premise)、主张(Claim)和推理规则, 论辩可认为是通过推理规则的作用使前提能够有效地支持主张的过程, 论辩挖掘的任务则是从数据中找到所有的主张、前提并建立它们之间的逻辑联系。论点结构与论辩过程是和研究与应用的目标密切相关的, 针对司法论辩、辩论或文本写作论辩可能有着不同的设计。

论辩挖掘是一个具有应用潜力的研究课题。随着电子政务和智能分析的推广以及社交媒体、论坛产生的用户生成数据不断增长, 从大规模信息流中发现、分离和分析论点的需求凸显了论辩挖掘的重要性, 同时为论辩挖掘提供了数据准备。目前, 用于从以上数据来源提取信息的技术主要基于统计和网络分析, 如舆论挖掘和社交网络分析。论辩挖掘系统可以对专业报纸文章、政府报告、法庭判决记录、在线社交网络内容中的论点、决策、评论等进行大

量的定性分析，为社会和政治科学领域的决策者和研究人员提供前所未有的自动化工具，为企业市场营销创造新的前景。

在本文中，我们将在论辩模型部分介绍理论上与实际中用于表示论点和论辩过程的基本元素；接着介绍从数据中挖掘结构化论点的基本方法；随后简述论辩挖掘技术在具体领域的应用；最后对论辩挖掘的挑战和前景做初步的总结和展望。

2 论辩模型

论辩在辩证法和哲学中有着悠久的根基，专门用于研究和分析陈述和断言如何被提出、辩论以及如何使观点之间的分歧得以解决并达成共识。除此之外，论辩已经渗透到语言、逻辑、修辞、法律、心理和计算机科学等多个学科领域。文献中关于论辩的理论模型是十分丰富的。最著名的是 Toulmin 模型[2]。Toulmin 提出，论辩的逻辑结构包括六个因素：主张(Claim)、事实材料(Grounds)、保证(Warrant)、模态词(Modality)、支援(Backing)和例外(Rebuttal)。主张指主观的观点或结论；事实材料可以是统计数据、常识和经验观察等，被用于支持主张；保证则可视为一种推理规则，确保事实材料能够使主张成立；模态词是指“可能”、“大概”等对主张的强度进行描述的词语；支援反映保证的权威性和有效性，表明保证的确具有证明力；例外则通常限定主张成立的条件或指出反例。图 1 给出了 Toulmin 模型中主要元素的一个示例。该示例展示了如何从事实材料出发、通过支援得到具有权威性的保证、排除例外、最终推出主张的过程。

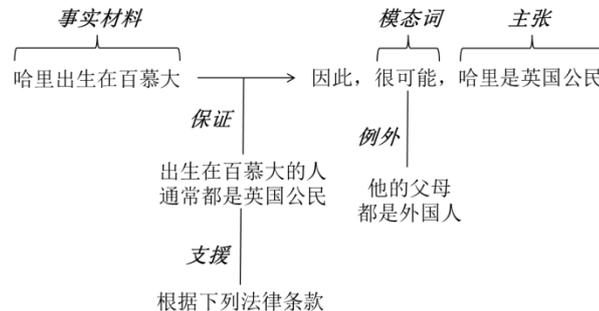


图1. Toulmin模型举例

另外一个著名模型是 Freeman 模型[3]。Freeman 模型认为论辩主要由前提(Premise)、结论(Conclusion)、模态词、反驳(Rebuttal)和反反驳(Counter-rebuttal)五个要素构成。Toulmin 模型和 Freeman 模型部分要素是彼此对应的，其他要素则体现出对过程中一些细节的不同理解，这里我们不再展开做深入的解读。通过对论辩模型的简单了解，我们就能感受到论证中体现出的结构性和过程性。

人工智能领域的论辩模型始于 Pollock[4]，Simari 和 Loui [5]以及 Dung[6]等开创性的工作并形成计算论辩这一个新的研究领域。该领域与知识表示、推理和多代理系统等方面研究有着紧密联系。Bentahar 等将论辩模型概括为三个主要类别：修辞、对话和单调模型[7]。修辞模型强调论辩过程中运用多种修辞手段对论辩话语的内容、形式进行策略性安排，试图使论辩话语兼具合理性和有效性；对话模型侧重于描述论点在对话结构中的连接方式，典型的场景是多个用户针对某一辩题进行辩论；单调模型则强调单人产生的文档或话语中论

点本身的结构和论点之间的关系等。无论哪类模型都具有一定的结构。

论辩挖掘通常基于结构化论辩模型，主要研究如何从自然语言文本中自动识别、比较和评估论辩，以满足信息化背景下人们对信息检索和信息抽取的更高需求。尽管 Toulmin 等理论模型已经对论辩结构进行了精彩刻画，在计算论辩中由于不同领域面对的具体问题不同，加之在技术上实现理论模型的难度较大，通常都会对模型进行一定的简化。Walton[8]给出了一个直观的论辩定义由三类部件(Component)构成：一系列前提，一个结论以及从前提到结论的推论。在文献中，结论(Conclusion)、主张(Claim)、中心(Thesis)通常都代表类似的角色，前提也常被称为证据(Evidence)或原因(Reason)，与 Toulmin 模型中的事实材料起到相同的作用。

在自然语言文本中进行论辩挖掘可根据论辩的结构特点划分为几个基本任务：论点部件的识别，其子任务包括主张检测(Claim Detection)和证据检测(Evidence Detection)；关系预测或结构预测，目的在于识别论点内部和论点之间的关系。具体的关系类型与任务有关，既可以是简单的相关关系，也可能细分为支持(Support)或攻击(Attack)关系等。

3 论辩挖掘方法

论辩挖掘系统需要解决一系列内在相关的子任务。图2展示了当前典型的论辩挖掘系统的处理流程[9]。系统输入非结构化文本，最终产生有结构化标注的文本。文本中检测到的论点内部和有关联的论点之间将会被连接起来形成论点图。下面介绍2个典型的子任务：论点部件检测和论辩结构预测。

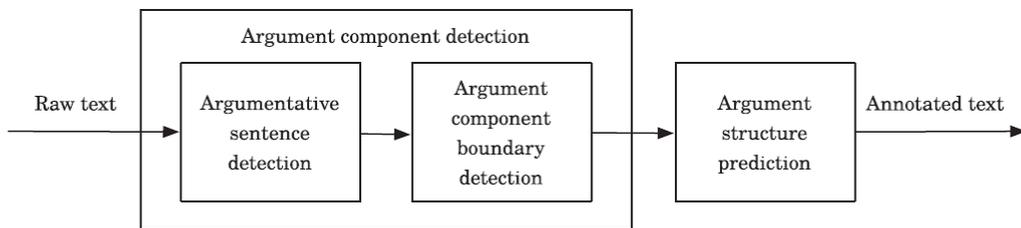


图2 典型论辩挖掘系统的处理流程

3.1 论点部件检测

论辩挖掘系统第一阶段目标主要是在输入的文本档内检测论点和论点部件。论点部件即论点内部的不同构成，如主张和前提。这个问题通常被划分为两个子问题来解决：提取有论辩性的语句和检测部件边界。前者通常在句子层次进行，后者则需考虑不同的粒度。

论辩句子检测：提取输入文档中包含论点的句子。该问题可以转化为一个分类任务，原则上可以使用任何类型的机器学习分类器来解决。然而，如何转化则可以有多多个不同解决方案。例如：

- (1) 训练二元分类器以区分论辩性语句和非论辩性语句，将识别论点具体部件的任务留给下一阶段；
- (2) 训练多类分类器来对所采用的论辩模型中存在的所有论点部件进行分类。

无论采用哪种方案都需要选择一种分类器。现有系统已经使用了各种各样的经典机器学习算法，但并没有明显的证据能证明哪个模型具有绝对优势。因此，迄今为止提出的方法通

常依赖于简单和快速的分类器，而将精力更多地投入到特征设计中。

常用的特征主要包括：(1)经典的词汇词袋特征：传统文本分类中的特征选择或权重估计策略均可采用；(2)基于本体、主题词表和词汇数据库等知识特征来缓解数据稀疏问题[10]；

(3)标点、情态动词和动词时态等功能信息；(4)通过外部分类器获得的复杂特征，如主观性评分、情绪分类器或命名实体识别系统等[10, 11]；(5)句间关系特征，利用RST或PDTB风格的篇章分析器构建特征[12]；(6)上下文特征，许多方法大量使用特定应用领域的知识，例如在法律文件中可借助于司法语言中频繁使用的特定句法描述符或关键短语，为相应结构的检测提供可靠的线索；(7)句法结构特征：尽管上下文信息很有效，但是它领域依赖的特点限制了论辩挖掘系统的泛化能力。有工作提出句法结构这种上下文无关信息作为特征。Lippi和Torrioni[13]使用Tree Kernel自动构建句法解析树的结构特征用于主张检测，其假设为句法树能够通过句子的修辞结构捕获预示主张存在的陈述语气。

论点部件边界检测：确定每个论点部件的确切边界，也称为论辩话语单元(argumentative discourse unit)识别。任务是在每个已经被预测为论辩性的句子中确定论点部件开始和结束的位置。这一问题的提出主要和对论点部件的粒度的要求有关，因为整个句子可能不完全对应于单个论点部件。有时一句话中可能包含多个论点部件，如既有主张又有前提；有时一个论点部件可能跨越多个语句，如用于支持主张的多个前提可能分布在多个句子中。这种粒度的要求通常和任务相关，也要考虑问题的复杂性。很多研究并不考虑边界检测问题而直接使用子句或整句作为处理单元或者假设论证话语单元已经被标注完备而关注于在此基础之上的其他任务。

典型的方法是将边界检测问题视为分割问题或序列标注问题，采用类似于处理序列标注的策略来加以解决。但整体来说，准确、自动的论点部件边界识别是比较困难的。

3.2 论辩结构预测

论辩结构检测旨在论点部件与边界检测的基础上，确定论点之间或论点部件之间的关联。这种语义上关联关系的识别往往涉及高级的知识表示和推理，因此是很有挑战性的。但论辩结构预测也是最能体现论辩挖掘特点的地方，有了这一步才能得到论辩挖掘的结构化输出。

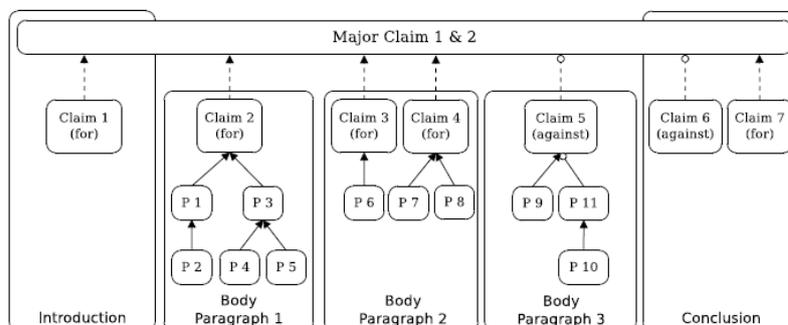


图3 议论文中论辩结构示意图

论辩结构预测的输出是连接检测出的论点及论点部件形成的图。图中的边可以表示蕴含、支持或攻击等不同关系。图3给出了议论文论辩结构的一个示意图[14]。显然，这种结构化

的表示更清晰地表达了文章的内部结构和逻辑关系，形成对文章的一种深度理解。这种表示对论证质量的判定、论据的获取与搜索等任务将是非常有帮助的。

Mochales Palau 等通过解析使用手动构建的上下文无关语法来预测论点部件之间的关系[15]。语法规则遵循司法文本中句子的典型修辞和结构模式。因此，这是一个高度类型特定的方法。Stab 和 Gurevych 的工作提出了一个经典机器学习解决方案，通过使用二元 SVM 分类器预测主张/前提模型中的链接并判断链接的类型是支持还是攻击[12]。Cabrio 和 Villata 借助于文本蕴含来推断论点间具有支持关系或是攻击关系[16]。Andreas Peldszus 等将结构预测问题划分为：关系识别、中心论点识别、功能识别等几个子任务并将这些任务的预测结果作为特征值计算论点之间关系形成图并采用最小生成树算法得到整体结构[17]。

4 论辩挖掘的应用

理想情况下，论辩挖掘得到结构化的论点，其中蕴含的逻辑关系对推理有着重要作用，因此论辩挖掘在特定领域有着诱人的应用前景。

司法领域：司法领域是率先关注论辩挖掘，也是论辩挖掘最成功的应用领域之一。欧洲人权法院通过从结构化法律文本集合中提取主张和支持主张的前提构建了 AraucariaDB 数据库[15]。这是一个高度专业化的、完整的论辩挖掘实践。通过论辩挖掘得到的数据库对于相似案件检索、自动司法判案等问题有重要作用。

医学领域：针对生物学和医学文本的数据集进行论辩挖掘吸引越来越多的关注。典型的任务包括建立症状描述和疾病之间的关联或基因和疾病之间的关联，这体现出论辩挖掘追本溯源的特点。构建的数据库可应用于自动医疗诊断和个性化的药物处方处理等方面。这方面一个典型任务是将科技文献中的句子分为介绍、方法、结果和讨论等类别，称为论证区域划分(Argumentative Zoning)[18]。

人文与教育领域：修辞、哲学和议论性文章构成了论辩挖掘关注的另一组研究对象。Lawrence 等人利用论辩挖掘技术对 19 世纪哲学论文集合中的论辩元素进行标注[19]。随着论辩挖掘的研究在计算语言学引起关注，相关的研究成果也被使用到教育领域，用来对学生议论文进行自动评分。Henning 等使用论辩挖掘得到的论点和论据的信息，对学生议论文的论辩结构进行标注，并使用该特征对学生议论文进行评分[20]。部分工作专注于识别议论文中的篇章角色，如导论、主旨、主要观点、事实论据、理论论据和结论等[21]，其基本目的也是建立起篇章级别的论辩结构并利用这种结构支持信息抽取与检索，如通过论据的主题标注与推荐支持议论文的写作[22]。

用户生成内容：社交媒体和互联网提供了多种包含论点或辩论的用户生成文本，如微博与论坛中的辩论，产品评论中对产品质量的评价与评价原因等。Ivan 和 Gurevych 从网络辩论论坛收集了用户的辩论内容，通过网络众包的方式标注了一个包含 16000 个论点对的数据集合。每一个论点对包括两个针对同一话题的论点，标签保证首个论点优于第二个[23]。同样是在辩论语料集合上，复旦大学团队对参与辩论双方的行为进行了标注并初步探索不同辩论行为和论点质量之间的关系[24]。人工标注论点质量费时费力，而且标注者难以形成统一意见，复旦大学团队的另一个工作即是使用论坛的自动标注信息作为论点质量的标签，并尝

试对论坛中的评论根据说服力进行排序[25]。在用户生成数据的洪流中，论辩挖掘将无结构文本转化为结构化内容的能力将为各种新的有趣应用提供可能。

5 结语

论辩挖掘任务与文本分割、情感分析、篇章结构分析等任务有着密切关联。论辩挖掘处理无结构文本输出有结构论点，其目的更加侧重于寻找结论所对应的原因，而不仅仅是意见或情绪。在这个意义上，论辩挖掘的关键目标是要比意见挖掘和情绪分析更进一步，与篇章结构分析相比有着更加明确的目的和应用倾向。

目前，论辩挖掘仍然面临很大挑战。首要的就是数据问题。通过机器学习和人工智能技术方法的任何论辩挖掘尝试都需要收集有标注的语料库用于训练和测试分类器。标注语料库通常本身是一个复杂和耗时的任务，对于论辩挖掘这一领域尤其困难，因为即使对人来说，论点部件的识别、它们的确切边界以及它们如何相互关联也是相当复杂的并且有争议的。本文中提到的很多参考文献都是基于自主构建的小规模数据集，其中多数数据集都是针对特定的目标或特定领域来构建的。数据的限制使得目前火爆的深度学习在论辩挖掘中并没有得到大规模应用，如何有效、快速地采集数据或研究无监督与半监督的方式进行论辩挖掘是值得探讨的问题。

展望未来，高性能的论辩挖掘工具的发展能够为社会科学和人文学科以及生命科学和工程领域的许多令人兴奋的新应用铺平道路。

参考文献

- [1] Trevor J. M. Bench-Capon and Paul E. Dunne. 2007. Argumentation in artificial intelligence. *Artificial Intelligence* 171, 10–15 (2007), 619–641.
- [2] Stephen Edelston Toulmin. 2003. *The Uses of Argument*. Cambridge University Press.
- [3] James B. Freeman. 1991. *Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure*. Vol. 10. Walter de Gruyter.
- [4] John L. Pollock. 1987. Defeasible reasoning. *Cognitive Science* 11, 4 (1987), 481–518.
- [5] Guillermo R. Simari and Ronald P. Loui. 1992. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence* 53, 23 (1992), 125–157.
- [6] Phan Minh Dung. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77, 2 (1995), 321–358.
- [7] Jamal Bentahar, Bernard Moulin, and Micheline B´elanger. 2010. A taxonomy of argumentation models used for knowledge representation. *Artificial Intelligence Review* 33, 3 (2010), 211–259.
- [8] Douglas Walton. 2009. Argumentation theory: A very short introduction. In *Argumentation in Artificial Intelligence*, Springer, 1–22.
- [9] Lippi, Marco and Paolo Torroni. 2016. Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology*, 16(2):10:1–10:25.
- [10] Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In *COLING 2014*, 1489–1500.

- [11] Ruty Rinott, Lena Dankin, Carlos Alzate Perez, Mitesh M. Khapra, Ehud Aharoni, and Noam Slonim. 2015. Show me your evidence—An automatic method for context dependent evidence detection. In EMNLP 2015, 440–450.
- [12] Christian Stab and Iryna Gurevych. 2014a. Identifying argumentative discourse structures in persuasive essays. In EMNLP 2014, 46–56.
- [13] Marco Lippi and Paolo Torroni. 2015. Context-independent claim detection for argument mining. In IJCAI 2015, 185–191.
- [14] Stab, Gurevych I. 2016. Parsing Argumentation Structures in Persuasive Essays[J]. arXiv preprint arXiv:1604.07370, 2016.
- [15] Raquel Mochales Palau and Marie-Francine Moens. 2011. Argumentation mining. *Artificial Intelligence and Law* 19, 1 (2011), 1–22.
- [16] Elena Cabrio and Serena Villata. 2012. Combining textual entailment and argumentation theory for supporting online debates interactions. In ACL, 2012, 208–212.
- [17] Peldszus A, Stede M. 2015. Joint prediction in MST-style discourse parsing for argumentation mining. In EMNLP 2015: 938-948.
- [18] Teufel, Simone. Argumentative zoning: Information extraction from scientific text. Diss. University of Edinburgh, 2000.
- [19] John Lawrence, Chris Reed, Colin Allen, Simon McAlister, and Andrew Ravenscroft. 2014. Mining arguments from 19th century philosophical texts using topic based modelling. In 1st Workshop on Argumentation Mining. ACL, 79–87.
- [20] Henning Wachsmuth, Khalid Al-Khatib, and Benno Stein, “Using Argument Mining to Assess the Argumentation Quality of Essays”, In COLING 2016, 1680-1692
- [21] Wei Song, Ruiji Fu, Lizhen Liu, Ting Liu. Discourse Element Identification in Student Essays based on Global and Local Cohesion. In EMNLP 2015, 2255–2261.
- [22] Wei Song, Ruiji Fu, Lizheng Liu, Hanshi Wang, Ting Liu. Anecdote Recognition and Recommendation. In COLING 2016, 2592-2602.
- [23] Ivan Habernal and Iryna Gurevych, “Which argument is more convincing? Analyzing and predicting convincingness of Web arguments using bidirectional LSTM”, In ACL 2016 1589-1599.
- [24] Zhongyu Wei, Yandi Xia, Chen Li, Yang Liu, Zach Stallbohm, Yi Li, Yang Jin, A Preliminary Study of Disputation Behavior in Online Debating Forum, In 3rd Workshop on Argumentation Mining. In ACL 2016, 166 - 171.
- [25] Zhongyu Wei, Yang Liu, and Yi Li. "Is This Post Persuasive? Ranking Argumentative Comments in the Online Forum.", In ACL(2) 2016, 195–200.