

## 平行推理机制:一种隐式篇章关系检测方法<sup>\*</sup>

洪宇<sup>1</sup>, 严为绒<sup>1</sup>, 车婷婷<sup>1</sup>, 梁颖红<sup>2</sup>, 姚建民<sup>1</sup>, 朱巧明<sup>1</sup>, 周国栋<sup>1</sup>

<sup>1</sup>(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

<sup>2</sup>(苏州职业大学 计算机系, 江苏 苏州 215104)

通讯作者: 洪宇, E-mail: tianxianer@gmail.com, http://nlp.suda.edu.cn/~hong/

**摘要:** 篇章是论元经过语义关联和结构化组织形成的自然语言文体。篇章分析研究的核心任务之一是解释论元的语义关系, 其中, 显式关系因具有直观线索而易于检测, 目前检测精度高达 90%; 相对而言, 隐式关系因缺乏直观线索而难于检测, 目前精度仅约 40%。针对这一问题, 基于一种“论元平行则关系平行”的假设, 并利用显式篇章关系易于检测的特点, 通过平行论元的识别与平行关系的消歧, 实现了一种显式关系平行推理隐式关系的隐式篇章关系检测方法。利用标准宾州篇章关系树库(Penn discourse TreeBank, 简称 PDTB)对这一检测方法进行评测, 结果显示, 精确率提升达 17.26%。

**关键词:** 篇章; 隐式篇章关系; 关系检测; 平行线索; 篇章关系消歧

**中图法分类号:** TP391

中文引用格式: 洪宇, 严为绒, 车婷婷, 梁颖红, 姚建民, 朱巧明, 周国栋. 平行推理机制: 一种隐式篇章关系检测方法. 软件学报, 2014, 25(11): 2528-2555. <http://www.jos.org.cn/1000-9825/4546.htm>

英文引用格式: Hong Y, Yan WR, Che TT, Liang YH, Yao JM, Zhu QM, Zhou GD. Parallel inference: A method of implicit discourse relation detection. Ruan Jian Xue Bao/Journal of Software, 2014, 25(11): 2528-2555 (in Chinese). <http://www.jos.org.cn/1000-9825/4546.htm>

### Parallel Inference: A Method of Implicit Discourse Relation Detection

HONG Yu<sup>1</sup>, YAN Wei-Rong<sup>1</sup>, CHE Ting-Ting<sup>1</sup>, LIANG Ying-Hong<sup>2</sup>, YAO Jian-Min<sup>1</sup>, ZHU Qiao-Ming<sup>1</sup>, ZHOU Guo-Dong<sup>1</sup>

<sup>1</sup>(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

<sup>2</sup>(Department of Computer, Soochow Vocational University, Suzhou 215104, China)

Corresponding author: HONG Yu, E-mail: tianxianer@gmail.com, <http://nlp.suda.edu.cn/~hong/>

**Abstract:** Discourse is a literary form that consists of semantically-related and well-structured arguments. One of the key tasks of discourse analysis is to resolve semantic relationship between arguments. Explicit relation is easy to detect with an accuracy of nearly 90% because of its direct cues. In contrast, implicit relation is difficult to detect with only an accuracy of nearly 40% since it has no direct cues. To solve the problem, paper proposes a hypothesis that parallel arguments normally have the consistent semantic relations. Based on the hypothesis and by utilizing the characteristics that explicit relation is easy to detect, the paper implements a method of implicit relation detection which uses explicit relation to infer implicit relation among parallel arguments. We evaluate the method on the standard penn discourse Treebank (PDTB). The experimental results show an improvement of 17.26%.

**Key words:** discourse; implicit discourse relation; relation detection; parallel cue; discourse relation disambiguation

篇章关系检测(discourse relation detection)是一种对论元之间语义关系进行自动识别与判定的自然语言处理技术, 对于篇章生成过程中的语义和语用规律分析, 以及借助语义关系的事件关系网络挖掘、语义计算和语

\* 基金项目: 国家自然科学基金(61373097, 61272259, 61272260, 90920004); 教育部高等学校博士学科点专项科研基金(20093201110006); 苏州市国际科技合作计划(SH201212)

收稿时间: 2013-05-14; 修改时间: 2013-09-02; 定稿时间: 2013-12-04

义搜索等应用研究,都具有重要的辅助作用。

宾州篇章关系树库(Penn discourse TreeBank,简称 PDTB)<sup>[1,2]</sup>对论元及其语义关系给出了较为完整的定义。其中,论元(argument,简称 Arg)定义为一种具有独立语义的文字片断,其区别于语义角色分析领域中论元的概念(在该领域中,论元是一种谓词受事结构中的名词性对象),PDTB 中的论元特指形成篇章关系的基本表现单元,如文字块、子句和句子等。论元关系则是论元在篇章结构中语义关联的性质,包含对比(comparison)、偶然(contingency)、时序(temporal)和扩展(expansion)这 4 大类,并且每类关系可根据语义进一步细化为多层次关系,比如,“偶然”关系可细分为因果(cause)和条件(condition)等子关系。

篇章关系检测系统的基本输入样本为特定篇章中的两两论元组合(下文简称为论元对),相应地,输出为论元对的语义关系标签(relation sense),如例 1 所示。

例 1:论元 1(Arg1):He worked all night yesterday (译文:他昨天工作了一晚上)  
 论元 2(Arg2):【So】he slept all day today (译文:【所以】他今天睡了一整天)  
 Relation Sense=Contingency.Cause (译文:关系标签=“偶然.因果”关系)

其中,根据论元对之间是否存在显式连接词,篇章关系可以划分为显式关系和隐式关系两类。比如,上例中,论元 1(Arg1)和论元 2(Arg2)之间存在显式连接词“so”(译文:“所以”),从而,论元之间构成显式篇章关系;否则,形成隐式篇章关系。相应地,篇章关系检测任务也划分为显式关系检测(explicit relation detection)和隐式关系检测(implicit relation detection)两项子任务。

目前,显式篇章关系检测性能已达到实用标准,准确率高于 90%<sup>[3]</sup>。相对来说,隐式篇章关系检测性能较低,准确率仅约 40%<sup>[4,5]</sup>。其原因在于,显式篇章关系中的连接词能够直白地表述论元之间的语义关系,比如 PDTB 关系样本库中,连接词“Because”(译文:“因为”)指向“Contingency.Cause”关系(译文:“偶然.因果”关系)的概率为 100%,从而以连接词为主要线索,借助连接词与篇章关系的一一映射,能够较为精确地推断特定篇章关系。与此相反,隐式篇章关系不具备显式连接词,换言之,缺乏指向特定关系的直观语言线索,仅能借助论元内容层面的语义关联性以及客观世界中事物的本源逻辑关系进行间接判断;尤其在语境中,两两论元往往具有多种不同的篇章关系,如果缺失显式线索的约束,则两者的关系存在一定的模糊性(即关系歧义)。由此,隐式篇章关系检测的难度较高,现有检测方法也因缺少关联线索的挖掘与应用,难以有效地提高检测性能。

针对上述问题,本文提出“论元平行则关系平行”这一基础假设(下文简称为平行假设),并基于这一假设,利用显式篇章关系易于检测的特点(精度高于 90%),形成一套“由显式关系推理隐式关系”的平行推理机制。例 2 给出了这一机制的样例,其中,论元对 X 和论元对 Y 具有平行语义。根据平行假设,两项论元对的关系也平行,即,论元对 X 与 Y 蕴含相同关系,从而当已知 X 的显式篇章关系为“偶然.因果”关系时,可平行地推理 Y 的隐式关系也为“偶然.因果”。

例 2:ArgPair\_X: McCain is beat by Bush (Arg\_X1), 【Connective=So】 Bush win the election (Arg\_X2)  
 译文:(显式论元对 X:麦凯恩被布什击败(论元 X1),【连接词=“所以”】布什赢得选举(论元 X2))  
 ArgPair\_Y: Bush beats McCain (Arg\_Y1), 【Connective=?】 winning the election (Arg\_Y2)  
 译文:(隐式论元对 Y:布什击败麦凯恩(论元 Y1),【连接词=未知】赢得了这次选举(论元 Y2))

基于平行假设的平行推理机制,能够有效地利用大规模数据中针对特定语义及其关系的多样化表述形式,将其中的显示关系作为参照物,分析“同指但不同质”的隐式关系,并以论元平行性为桥梁,形成一种“显式指导隐式”的机器学习模式。但利用这一模式完成隐式篇章关系的判定,还需重点解决如下问题:

- (1) 如何挖掘有效的平行论元对,辅助以论元平行性为桥梁的关系推理?
- (2) 促成论元对平行的特征并不单一,不同特征往往反映不同关系,形成歧义。如何消除歧义?

针对上述问题,本文相应地提出检索驱动平行论元对挖掘方法、基于平行概率模型的关系推理方法以及基于同指焦点的歧义关系屏蔽方法,并形成了一套基于平行推理机制的隐式篇章关系检测方法。利用国际标准篇章关系分析树库(PDTB)对该方法进行评测,检测精度优于现有方法约 17.26%。

本文第 1 节回顾篇章关系检测的相关研究。第 2 节详述篇章关系检测的任务体系。第 3 节给出本文方法的

设计动机与基本思想,即平行假设思想与推理机制.第 4 节介绍基于平行推理机制的隐式篇章关系检测方法框架、方法模块及数据处理流程.第 5 节详细介绍检索驱动的平行论元对挖掘方法.第 6 节介绍面向关系推理的平行概率模型.第 7 节给出基于同指焦点的歧义关系识别与屏蔽方法.第 8 节介绍实验设计.第 9 节给出实验结果及相关分析.第 10 节总结全文.

## 1 相关研究

本节针对篇章关系检测领域的现有研究给予概述,包括基于语言特征的关系分类方法、基于概率统计的关系推理方法以及篇章关系消歧的前瞻性尝试,然后给出小结.

- 基于语言特征的篇章关系分类

篇章关系检测的主体研究集中于基于关系特征的机器学习方法,其中,尤以语言学特征的分析与应用以及依赖语言学特征的关系分类方法最为热门.其中,Pitler 等人<sup>[6]</sup>利用不同类别的篇章关系样本训练分类器,形成有监督的关系分类系统,并重点测试了情感极性、动词类型、动词短语长度、词形和上下文等特征对关系分类的影响.这一系统在 PDTB 测试集上获得了优于随机关系分类的基本性能.

Lin 等人<sup>[7]</sup>继承了 Pitler 等人<sup>[6]</sup>的方法框架,即有监督的关系分类系统,并在这一基础上细化了上下文特征的采集技术,且利用上下文作为论元之间衍生特定语义关系的外部特征;同时,结合使用了 Soricut 等人<sup>[8]</sup>提出的论元内部结构特征与依存特征,由此形成的分类系统应用于隐式关系分类,在 PDTB 测试集上获得了 40.2%的精确率.随后,Wang 等人<sup>[4]</sup>利用树核方法提升了句法结构特征的区分能力,但这一特征的优化并未给予显著的性能提升(约 40.0%的精确率),仅略优于以浅层句法树为特征的关系分类性能.

- 基于概率统计的篇章关系推理

概率统计在分析语言学规律的可能性和置信性方面有着长期深远的工作,而直接应用于篇章关系分析的相关工作则起始于 Marcu 等人<sup>[3]</sup>.Marcu 等人的观点是:两两共现的词特征能够在一定程度上反映两两共现的论元关系.比如,词特征“杀戮”与“伤亡”往往反映一种事实性的因果关系,从而,分别蕴含“杀戮”与“伤亡”的毗连论元之间也应具有这种语义关系,或者具有这种语义关系的概率极高.从而,Marcu 等人借助训练语料获取了词特征共现概率与篇章关系概率的对应性,形成了一套基于概率统计的篇章关系推理机制.相对于推理显式关系,Marcu 等人的方法更适于推理隐式关系,其原因在于:与共现词特征相比,显式连接词指向关系类型的概率更为精确.

因此,后续基于概率统计的篇章关系研究逐渐将目标转向隐式关系,并于 2006 年之后使隐式关系检测问题成为新的研究热点.其中,Saito 等人<sup>[9]</sup>继承了 Marcu 等人<sup>[3]</sup>的推理机制,并专门尝试利用语言学单位的共现概率以及共现单位对特定篇章关系的表述能力进行隐式篇章关系的检测,其主要研究成果是:验证了共现短语及其指向篇章关系的概率能够更精确地辅助关系推断.随后,Zhou 等人<sup>[5]</sup>验证了三元文法的共现概率及其关系指向,也能提高隐式关系的判定.相对而言,Zhou 等人比 Saito 等人更强调语序,但不局限于严格的语法规范,从而在一定程度上同时满足了词特征相互组合的连贯性和灵活性,由此产生的有效共现文法与篇章关系的映射样本得以扩展,也即增加了有效的推理线索.

- 篇章关系消歧

篇章关系包含两种层面的歧义:其一是显式关系歧义,源自连接论元的显式连接词本身具有多义性<sup>[10]</sup>,如 PDTB 中的连接词 since 既可表述时序关系(意为“自从”),又可表述因果关系(意为“由于”);其二是论元内容本身蕴含多种关系特征,且不同特征指向的关系都具有一定的合理性,从而一项论元对的关系可以获得多种解释,形成歧义关系,又因这一现象多发于隐式篇章关系,因此称为隐式关系歧义.比如,论元“少壮不努力”和论元“老大徒伤悲”中,“少壮”与“老大”之间形成了对比关系解释,而“不努力”与“徒伤悲”之间形成了因果关系解释,这一模棱两可的关系解释即为隐式关系歧义.

截至目前,专门针对篇章关系歧义的研究尚不全面,尤其是针对隐式关系歧义的研究尚为空白,仅有 Miltakaki 等人<sup>[3]</sup>和 Pitler 等人<sup>[11]</sup>针对多义连接词引起的显式关系歧义进行了局部研究.其中,Miltakaki 等人建

立了基于句法特征的最大熵模型,用于对 PDTB 中 since,while 和 when 这 3 种连接词的消歧处理.Pitler 等人扩充了消歧对象,但实验结果显示,将消歧用于显式篇章关系检测时,仅能提供约为 1%的性能改进.原因在于多义连接词数量不多,从而引发的歧义关系样本较少,即使全部正确处理也不能为显式篇章关系检测引入显著的性能提高.相对而言,Pitler 等人的错误分析验证了论元内容本身对关系判定的重要性,显然,论元内容中存在的隐式歧义特征是克服关系歧义的重要难点.

- 现有研究缺陷

显式篇章关系检测的现有性能已经趋近实用化(精度高于 90%),相对而言,隐式篇章关系性能较低(精度约 40%),因此,篇章分析领域的重点是提出行之有效的隐式关系检测手段.尽管上述概率统计方法为隐式篇章关系检测研究开辟了新途径,但获得的性能仍然偏低,其缺陷主要包括如下 3 点:

- (1) 上述方法皆为有监督的概率统计学习方法,共现特征与篇章关系的映射样本以及相应的概率指标都来源于固定且有限的语言学资源,无法获得所有映射样本,且映射样本的概率指标并不健壮.换言之,有效推理线索不足.
- (2) 采用句子中孤立的局部片断(包括词、短语和三元文法)推理论元(包括子句和句子)整体之间的语义关系,其本身即存在表义不全和表义不准的偏差,由此推断论元关系也不可避免地产生关系的偏差.换言之,推理线索不准.以共现词特征为例,虽然“杀戮”与“伤亡”往往表征因果关系,并且论元“伤亡数字持续攀升”和论元“杀戮并未停止”分别包含了上述特征,但两论元之间实际上表述了一种让步关系(在 PDTB 中,让步关系为对比关系的子关系),而不是因果关系;相应地,正确的语义理解也是“虽然伤亡数字仍持续攀升,但是杀戮并未停止”,而不是“由于伤亡数字持续攀升,所以杀戮并未停止”.
- (3) 论元内容中隐含的多种关系特征往往引入关系歧义,误导隐式关系推理.

因此,基于概率统计的隐式关系推理方法须在解决表义不全和表义不准的基础上,采用无监督的统计学习方法,实现在大规模的数据资源中获取更多有效推理线索,并屏蔽其中的歧义关系,才能获得实质性的检测性能提升.本文即针对上述问题,提出了基于平行推理机制的隐式篇章关系检测方法.

## 2 篇章关系检测任务定义

目前,篇章关系分析的对象是一对位置毗连的文字片断,它们既可以是一条句子中先后连续出现的两个子句,也可以是以句号为间隔的两个连续完整的句子.此外,标准篇章关系树库(PDTB)也给出了少量文字块(chunk)与子句的搭配.PDTB 统一将这类文字片断称为论元,意为具有独立语义的论述单元,而一对有待判定关系的论元样本,称为待检论元对.一项论元对中先出现的论元称为 Arg1,即前端论元(或简称论元 1),后出现的论元称为 Arg2,即后端论元(或简称论元 2).

篇章关系检测任务即是针对 Arg1 和 Arg2 之间的语义关系进行判定的任务.如图 1 所示,关系检测的输入为论元对 Arg1 和 Arg2,输出为 Arg1 和 Arg2 的关系类型标签(relation sense).

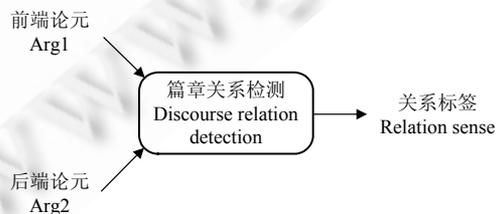


Fig.1 Frame of discourse relation detection task

图 1 篇章关系检测任务框架

PDTB 针对篇章关系检测任务给出了一套关系类型体系,共包含 39 种关系类型,分布于由宏观至具体的 3 个层次之中,见表 1.

**Table 1** PDTB discourse relation types in different levels (Level 1~Level 3)

表 1 PDTB 中的篇章关系类型(第 1 层~第 3 层)

Level 1(第 1 层)	Level 2(第 2 层)	Level 3(第 3 层)
Temporal(时序)	Synchronous(同步)	-
	Asynchronous(异步)	Precedence(优先) Succession(后续)
Comparison(比较)	Contrast(对比)	Juxtaposition(并列) Opposition(转折)
	Pragmatic contrast(语用对比)	-
	Concession(让步)	Expectation(期望) Contra-Expectation(相悖期望)
	Pragmatic concession(语用让步)	-
Contingency(偶然)	Cause(因果)	Reason(原因) Result(结果)
	Pragmatic cause(语用因果)	-
	Condition(条件)	Hypothetical(假设条件) General(综合条件) Unreal present(当前假设条件) Unreal past(以往假设条件) Factual present(当前真实条件) Factual past(以往真实条件)
	Pragmatic conditon(语用条件)	-
Expansion(扩展)	Exception(排除)	-
	List(枚举)	-
	Conjunction(联合)	-
	Instantiation(实例化)	-
	Restatement(重述)	Generalization(概括重述) Specification(详细重述) Equivalence(同义重述)
	Alternative(取舍)	Conjunctive(联合取舍) Disjunctive(孤立取舍) Chosen alternative(两者择一)

篇章关系检测任务包含两项子任务:显式关系检测和隐式关系检测.从任务定义的角度观察,两项子任务的唯一区别是输入端的论元对是否具有显式连接词(如“Therefore”,译文:“因此”);PDTB 指南(<http://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf>)共给出了 100 个英文显式连接词,具有显式连接词则为显式关系检测,不具有显式连接词则为隐式关系检测.

从任务难易的角度分析,两项子任务的区别较为明显:显式关系的连接词存在于论元对原文中,且其词义本身可以直观地反映关系类型,如显式连接词“因为”能够直观地反映论元对的关系为“偶然.因果”类型,在忽略连接词歧义(如“Since”,译文:“自从”或“由于”,既反映“时序.异步”关系,又反映“偶然.因果”关系)的情况下,甚至无需深入理解论元对的语义即可给予正确输出,现有显式篇章关系检测的精准性能(精度高于 90%)即折射了这一现象;相对来说,隐式关系的论元对原本文本地缺失显式连接词,即使 PDTB 语料资源的标注者,也需深入理解论元对的语义,且揣测其语义关系,并人工地推断连接词(PDTB 将人工推断的连接词称为隐式连接词或主观连接词),才能给出恰当的关系类型.从而,借助机器学习实现隐式篇章关系检测的难度较高,现有性能仅约 40%.本文研究的对象即为隐式篇章关系检测.

### 3 面向隐式关系检测的平行推理机制

针对隐式篇章关系检测子任务,提出了一种基于平行假设的平行推理机制,并利用显式关系易于判定的特点,进一步形成了显式指导隐式的平行推理机制.本节分别给出基于平行假设和显式指导隐式的平行推理思想.

### 3.1 平行假设思想

平行假设的核心思想是:如果一对论元与另一对论元具有一致语义(即论元语义平行),则两对论元的篇章关系也一致(即论元关系平行).详细而言,给定论元对  $X$  与论元对  $Y$ ,如图 2 所示,两者的语义平行必须保证  $X$  的  $ArgX1$  与  $Y$  的  $ArgY1$  平行,且  $X$  的  $ArgX2$  与  $Y$  的  $ArgY2$  也平行;相应地,关系平行是指“ $ArgX1$  和  $ArgX2$  的关系  $Relation-X$ ”与“ $ArgY1$  和  $ArgY2$  的关系  $Relation-Y$ ”具有一致性,即,论元对  $X$  与论元对  $Y$  具有相同的篇章关系.

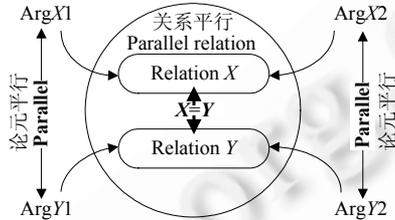


Fig.2 Theory of parallelism hypothesis

图 2 平行假设原理

根据上述平行假设的约束,论元平行性必须保证论元位置(前端或后端)的一一对应性.如图 2 所示,两者的前端论元  $ArgX1$  与  $ArgY1$  具有一致语义,后端论元同理,而不是前、后端交错平行.其原因在于论元的搭配顺序不同,反映的关系也可能不同.如例 3 所示,虽然论元相同,但采用不同搭配顺序后,产生的语义关系截然不同(例 3 中的关系先后为扩展和因果).从而,论元交错平行往往并不能保证关系平行,容易对“论元平行则关系平行”的推理机制产生误导.

例 3: I ate the apple ( $Arg1$ ), **【and】** I felt it was good ( $Arg2$ ) (我吃了这个苹果,【并且】觉得味道不错)

I felt the apple was good ( $Arg1$ ), **【so】** I ate it ( $Arg2$ ) (我觉得这苹果味道不错,【所以】把它吃了)

在平行假设中,论元的语义平行性类似于跨语言信息检索<sup>[12]</sup>和机器翻译<sup>[13]</sup>领域中的双语平行性,包括但并非强调用词、语法、句法和依存等特征的绝对一致性(同质同指),而是更多地强调论元表达的含义具有一致性(异质同指).例 4 给出了双语平行与单语平行的对比样例.

例 4: 双语平行样例:“Fall in love”平行于“坠入爱河”(注:非字面直译,即异质同指)

单语平行样例:“Obama won the election”平行于“Obama won the White House”(注:非字面平行) (译文:“奥巴马赢得选举”) (译文:“奥巴马入主白宫”)

平行语义学的一项重要应用是为面向语言分析与理解的有监督(supervised)或半监督(semi-supervised)机器学习提供标准训练资源,其优势在于这类平行资源能为分析待测样本的语言现象提供标准的参考或类比对象.比如,在机器翻译领域中,利用双语平行资源提取的翻译规则,可用于优化候选译文的调序;而在句法分析领域中,单语平行资源蕴含的平行句法树,可用于优化句法消歧.受这类平行语义应用技术的启示,本文尝试利用语义平行性实现平行论元之间的语义关系类比,并借此在平行论元对之间以显式关系为参考对象,实现隐式关系的自动推理.

### 3.2 显式指导隐式的平行推理机制

根据平行假设,可以形成如下平行推理机制:如果两项论元对的语义相互平行,且已知其中一项论元对的篇章关系,则可将这一论元对的篇章关系平行地指定为另一论元对的篇章关系,形成“由已知关系推理未知关系”的机制,即平行推理机制.但是,这一平行推理需要先验地已知其中一项论元对的篇章关系,从而形成由已知关系推理未知关系的机制.

针对平行推理机制的上述先验条件,本文根据篇章关系分析中显式关系易于判定(精度高于 90%)而隐式关系难于判定(精度仅约 40%)的特点,将显式关系作为平行推理过程中的近似已知条件,对隐式篇章关系予以推理.详细而言,如果两项论元对的语义相互平行,且已知其中一项论元对具有显式篇章关系,则可利用现有显式

篇章关系的检测方法,对这一论元予以先期的关系认定,并将其作为已知条件,利用平行假设推理另一项论元对的隐式关系,从而形成显式指导隐式的平行推理机制。

比如,下述例 5 中的两项论元对具有平行语义,其中,包含连接词“所以”(即 **Connective=So**)的论元对具有显式关系,从而可以利用现有基于连接词特征的显式关系检测方法,通过连接词“所以”与“偶然.因果”关系的一一映射,先期判定这一论元对具有“偶然.因果”关系,并在此基础上,根据平行假设推理另一项论元对的隐式关系也为“偶然.因果”。

例 5:显式关系论元对:McCain is beat by Bush (**Arg1**). **【Connective=So】** Bush win the election (**Arg2**)

隐式关系论元对: Bush beats McCain (**Arg1**). **【Connective=?】** winning the election(**Arg2**)

## 4 基于平行推理的隐式关系检测框架与流程

### 4.1 隐式关系检测方法框架

本文基于上述平行假设及显式指导隐式的平行推理机制,尝试建立一种隐式篇章关系的自动检测方法.本节首先介绍方法的主要模块,然后给出处理流程.整体的方法框架如图 3 所示,主要包含如下 3 个模块:检索驱动的平行论元对挖掘、面向关系推理的平行概率模型和基于同指焦点的歧义关系屏蔽。

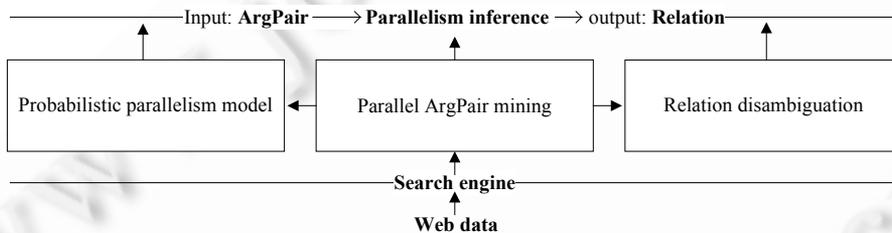


Fig.3 Frame of implicit relation detection based on parallelism inference

图 3 基于平行推理机制的隐式篇章关系检测框架

下面对各个模块的核心问题与研究内容进行详述。

#### • 平行论元对挖掘模块

平行论元对挖掘的目标是为待测论元对获取语义平行的其他论元对,借以辅助“论元平行则关系平行”的推理机制.同时,为了满足显式指导隐式的平行推理条件,挖掘过程不仅需要保证论元语义一致性,而且需要保证挖掘到的平行论元对具有显示关系(即包含显示连接词)。

与相关性或相似性相比,平行性往往对语义和语用的一致性有着更为严格的要求.从而,与相关论元或相似论元相比,符合平行性的论元较为稀疏.比如,同一篇章中的论元衍生于同一主题,两两之间具有不同层次的相关性,从而,相关论元易于查找且数据充分;而平行论元类似于论元的复述,绝大部分论元在同一篇章中复述的概率较低,往往离散于不同媒体和不同形式的篇章中,难以查找且单位数据集内的稀疏性较高.针对这一问题,本文采用一种检索驱动的平行论元对挖掘方法,利用搜索引擎 Google 在大规模 Web 信息中提取平行论元对,借以解决上述数据稀疏问题(方法细节如第 5 节)。

#### • 关系推理的平行概率模型

平行概率模型针对的主要问题是:若论元对的平行程度不同,则关系的平行概率不同.产生这一问题的原因在于:绝对的平行假设难以进行显式指导隐式的平行推理,而相对的平行假设虽然能够进行上述推理,但是语义平行的相对性必然引入平行程度的不同层次.显然,语义平行程度不同,语义关系的平行程度也不同,从而为平行推理机制引入概率问题.下面详述这一问题。

根据平行假设所述,两对语义相互平行的论元对往往具有平行(即相同)篇章关系.然而,绝对的平行假设仅仅建立在论元对复写(即重新抄录)这一理想情况下,即两对论元对字面完全一致.而这一理想情况的发生概率

较低,即使在大规模数据集中进行检索(排除“拷贝检测”<sup>[14]</sup>领域中的复制样本),往往也难以获得复写论元对(这一现象也恰恰反映了语言组织与应用的灵活性);并且,既然为复写论元对,那么复写与被复写的论元对在是否具有连接词方面也将具有绝对一致性,或都具有连接词(形成显式关系),或都无连接词(形成隐式关系),从而也无法支持显式指导隐式的平行关系推理。

比较而言,相对的平行假设侧重于两两论元对语义而不是字面的一致性,类似于异质同指的复述与重构。换言之,两两论元对陈述了相同含义即可。这类平行论元对的出现概率较高,且因为字面组成形式的一致性要求较为宽松,而各类媒体(如新闻和微博)、领域(如自然科学与生物医学)和个人(如主页和聊天记录)等信息源对同一语义的论述习惯和文法形式多种多样,使得平行论元对的显式关系(有连接词)和隐式关系(无连接词)并存的概率也较高,比如,新闻媒体的信息表述风格较为严肃,语义关系往往严格利用显式连接词予以明确表征;反之,微博的信息表述简洁且随意,缺省现象严重,语义关系论述往往不具有显式连接词。因此,基于相对平行假设,能够较好地利用大规模数据和检索技术,同时挖掘同一语义关系的显式和隐式样本,有效支持显式指导隐式的平行推理机制。

但是,相对的平行假设仅仅强调语义一致性,而语义一致性本身具有程度上的层次区别。比如,论元  $X$ “他难得读书”和论元  $Y$ “他从不读书”都与论元  $Z$ “他不努力读书”有相同的含义,但  $Y$  与  $Z$  的一致性略高。而语义一致程度的不同往往造成语义关系类型的不同,容易导致关系推理错误。

比如在例 6 中,给定待测论元对  $Z$  和两项挖掘所得的平行论元对  $X$  与  $Y$ ,观察可发现如下问题:虽然论元对  $X, Y$  和  $Z$  相互平行,但  $Y$  与  $Z$  的平行程度更高,适合将  $Z$  的关系认定为  $Y$  的关系,即“Contingency.Cause”(偶然.因果)关系,而不是  $X$  的“Comparison.Concession”(对比.让步)关系。

例 6:平行论元对  $X$ :他难得读书(Arg1),老师【却】批评他(Arg2) Relation Sense=Comparison.Concession

平行论元对  $Y$ :他从不读书(Arg1),【所以】老师批评了他(Arg2) Relation Sense=Contingency.Cause

待测论元对  $Z$ :他不努力读书(Arg1),【Connective=?】老师批评了他(Arg2) Relation Sense=?

从而,论元对语义平行程度的不同,引出了“论元平行则关系平行”的概率估计问题,即,根据语义平行程度估算关系平行的可能性问题。本文平行概率模型即为针对这一问题提出的解决方法,其核心问题是:对挖掘所得的平行论元对进行平行程度的度量,并按程度不同进行关系概率的计算(方法细节见第 6 节)。

#### • 歧义篇章关系屏蔽

平行推理机制的核心思想是利用平行论元对的显式关系推理待测论元对的隐式关系,其中,挖掘所得的平行论元对及其显式关系,是实现这一推理的先验知识。然而,论元对的关系存在多种解释,形成歧义,当歧义出现于作为先验知识的显式关系时,利用显式指导隐式的平行推理必然受到混淆,增加误判的几率。歧义篇章关系的产生原因主要包括如下两类:(1) 连接词歧义;(2) 焦点歧义。定义如下:

- 连接词歧义特指反映显式关系的连接词本身具有多种词义,且不同词义反映不同篇章关系类型。比如,连接词“Since”具有“自从”和“由于”两种词义,且前一词义反映时序关系,后一词义反映因果关系。从而,当显示关系由这一连接词触发时,平行推理机制无法确定其具有哪一种显式关系,导致显式指导隐式的推理陷入混淆。
- 焦点歧义特指论元对的内容存在多种可能的论述焦点,且不同焦点暗喻不同篇章关系类型。比如在例 7 中,论元对  $X, Y$  和  $Z$  相互平行,平行程度已达到复写标准,即绝对平行。尽管如此,当附着不同连接词后,表述的篇章关系却截然不同: $X$  表述“Contingency.Cause”(偶然.因果)关系; $Y$  表述“Comparison.Concession”(对比.让步)关系; $Z$  表述“Temporal.Asynchronous”(时序.异步)关系。对这一现象的合理解释即为论元对中暗喻关系的焦点不同,比如例 7 的论元对  $X$  中,暗喻关系的焦点(focus)为“工作了一夜”,强调因由; $Y$  中的焦点为“睡了一天”,强调前一论元的对立面; $Z$  中焦点为“昨天”和“今天”,强调时间顺序。

例 7: $X$ :他昨天工作了一夜(Arg1),【所以】今天睡了一整天(Arg2) Focus=“工作了一夜”

$Y$ :【虽然】他昨天工作了一夜(Arg1),【但是】今天睡了一整天(Arg2) Focus=“睡了一天”

Z:他昨天工作了一夜(Arg1),【随后】今天睡了一整天(Arg2) Focus=“昨天”+“今天”

显然,例 7 中的 3 种关系都具有一定合理性,都可以用于显式指导隐式的平行推理,而对于一项待测隐式论元对,其篇章关系往往唯一(PDTB 中,论元对含有唯一关系的样本比例约为 99%),从而,当挖掘所得的平行论元对集合同时包含上述多种关系时,显式指导隐式的平行推理将陷入难以取舍的局面。

相对而言,歧义连接词数量较少,且发生概率较低.比如,PDTB 给出的 100 项显式连接词中,具有明显歧义的词项不足 10%.表 2 摘录了 PDTB 中常见的连接词,并给出了每个连接词表征特定关系的最大概率,其中,明显具有歧义(最大概率接近 50%)的连接词为“since”等,仅为 4 项(如表 2 中下划线标示的常用连接词).从而,连接词歧义对平行推理的误导并不严重.不同的是,焦点歧义往往出现于论元内容中,频繁且难于识别.本文前期研究中,从 PDTB 随机抽取了 800 项论元对,通过人工标注与交叉检验(3 人交叉检验),发现 22%的论元对存在焦点歧义问题.因此,本文重点针对焦点歧义进行处理,并提出了基于同指焦点的隐式篇章关系消歧方法。

Table 2 Mapping table of frequent PDTB connectives and their most possible relations

表 2 PDTB 常见连接词及其最大概率关系映射表

Comparison	<u>While (66.07%)</u> , But (97.19%), yet (97.03%), still (98.42%), however (99.59%), although (99.70%), though (100.00%) ...
Expansion	In fact (92.68), indeed (95.19%), and (96.83%), or (96.94%), instead (97.32%), unless (98.95%), <u>also (99.94)</u> , for example (100.00%), in addition (100.00%), moreover (100.00%), for instance (100.00%), ...
Contingency	<u>Since (52.17%)</u> , if (95.99%), because (100.00%), so (100.00%), thus (100.00%), as a result (100.00%)...
Temporal	<u>meanwhile (48.70)</u> , <u>as (70.26%)</u> , when (80.18%), until (87.04%), then (93.24%), once (95.24%), later (98.90%), after (99.65), before (100.00%) ...

#### 4.2 基于平行推理机制的关系检测流程

基于平行假设和平行推理机制,针对任意一项隐式关系的待测论元对,本文通过基于 Web 数据的平行论元对挖掘、基于平行概率模型的关系推理和基于同指焦点的歧义关系屏蔽这 3 项主要技术,实现隐式篇章关系的自动检测,检测流程主要包含 3 个步骤,如图 4 所示。

Input: Implicit ArgPair  $X$ (输入待测论元对  $X$ ).

Method Body(方法体):

Step 1:抽取  $X$  的词特征构造查询  $Q$ ,利用搜索引擎,检索平行于  $X$  的论元对,

采集其中具有显式关系的论元对形成线索集合  $U_{key}$ ;

Step 2:基于线索集合  $U_{key}$ ,利用显式指导隐式的平行推理机制和关系平行概率模型,推断待测论元对  $X$  的隐式关系;

Step 3:基于大规模静态数据(本文采用 GIGAWORD),训练同指焦点概率,并用以屏蔽 Step 2 的歧义关系;

Output: Implicit Relation(输出隐式关系).

Fig.4 Basic detection procedure

图 4 基本检测流程

检测流程中的第 1 步(Step 1)利用检索技术先期挖掘大量关于待测论元对的 Web 数据,抽取其中具有显式关系的论元对,并通过检验平行程度,选取其中  $N$  项平行程度最高的论元对(参数  $N$  的取值如第 9 节实验部分),形成推理线索集合  $U_{key}$ ,这一步骤的关键问题是查询  $Q$  的构建方式和平行程度的合理计算;第 2 步(Step 2)利用先验统计所得的连接词与显式关系映射表(表 2 给出了部分映射关系),借助集合  $U_{key}$  中的显式关系论元对,实现显式指导隐式的关系推理,推理过程中,集合  $U_{key}$  内的每项论元对都作为推理线索,根据其待测论元对的语义平行程度,计算其关系平行概率,并利用平行概率模型推断集合  $U_{key}$  的全局平行概率,作为关系推理的数值依据,这一步骤的关键是平行概率模型的构建;第 3 步(Step 3)针对集合  $U_{key}$  中蕴含多种显式关系而造成推理混淆的问题,对  $U_{key}$  中的每项论元对进行同指焦点概率的计算,即度量  $U_{key}$  中每项论元对与待测论元对具有相同焦点的概率,并将概率较高者作为平行关系推理的主要参考对象,优化步骤 2(Step 2)中的关系平行概率模型,这一步骤的重点是同指焦点概率的计算方法.下面对各个步骤的技术细节加以介绍。

## 5 检索驱动平行论元对挖掘

检索驱动平行论元对挖掘是实现平行关系推理的基础,主要包括两个部分:其一,构建高质量查询关键词,并借助搜索引擎(本文采用 Google)检索相关于待测论元对的 Web 数据,抽取其中具有显式关系的论元对,形成平行论元对的候选集合;其二,从候选集中选择与待测论元对语义平行的样本,构成显式关联且语义平行的论元对集合(简称线索集).下面分别予以介绍.

### 5.1 查询构建

如第 2 节所述,论元对包括 Arg1(前置论元)和 Arg2(后置论元),而论元对的隐式关系是由 Arg1 和 Arg2 的内容关联产生的语义联系.由此,为了保证论元对整体参与语义平行计算(而不是局部的论元 Arg1 或 Arg2),且保证检索获得的相关信息极大概率地蕴含 Arg1 与 Arg2 的内在联系,查询构建过程需要同时抽取 Arg1 和 Arg2 的特征,并采用特征关联规则形成查询描述.

为此,本文分别抽取 Arg1 中的  $n$  元文法(称为前置文法,即 PreGram)与 Arg2 的  $n$  元文法(称为后置文法,即 PosGram),将两者同时作为查询特征,并利用特征关联规则“PreGram”+“PosGram”构造查询,如图 5 所示.这一规则中,双引号和加号是必须项,不能省略.其功能在于:当使用搜索引擎 Google 进行检索时,这种关联规则能够保证伪相关反馈(也称为检索结果)同时包含“PreGram”和“PosGram”两项特征,从而使“PreGram”和“PosGram”协同产生的论元对在伪相关反馈中高频出现,有益于发现和采集语义平行的论元对.此外,本文实验部分对不同  $n$  元文法( $n=1\sim 5$ )进行检验(如第 9 节),并验证二元文法(bigram)形成的“PreGram”和“PosGram”更易于挖掘平行论元对.由此,将 Arg1 与 Arg2 的二元文法作为查询特征的抽取对象,辅助平行论元对的挖掘过程.



Fig.5 Query representation

图 5 查询表示

基于上述查询构建方法,利用搜索引擎 Google 进行检索,可获得伪相关反馈列表(即检索结果列表).针对这一列表,挖掘过程遍历每一条检索结果的网页快照(snippet)和链接标题(title),并从中采集候选平行论元对.采集过程遵循如下规则:

- (1) “PreGram”和“PosGram”分别出现于两条不同的子句或句子内;
- (2) 包含“PreGram”和“PosGram”的两条子句或句子相互毗连;
- (3) 包含“PreGram”和“PosGram”的两条子句或句子由唯一一项显式连接词连接而成.

图 6 给出了候选平行论元对的采集样例,其中,给定的待测隐式论元对为“I hate you”(Arg1)和“I love you”(Arg2),选取的论元特征为二元文法“PreGram=hate you”和“PosGram=love you”,利用特征关联规则形成的查询为“hate you”+“love you”,利用伪反馈列表(图 6 仅给出伪反馈列表的局部片断)和上述采集规则,可采集到“ArgPair1: I hate you then I love you”和“ArgPair2: I hate you but I love you”等具有显式关系的平行论元对.这类平行论元对的显式关系(如上例中显式连接词“then”和“but”反映的关系)能够为推理待测论元对的隐式关系提供重要线索和参考.

在此基础上,本文将待测论元对中包含的所有二元文法 PreGram 和 PosGram 进行组合(如图 7 所示),形成多项查询,并利用上述检索和挖掘过程对所有查询的伪反馈列表进行候选平行论元对的采集,合并构成候选平行论元对集合,投入后续语义平行度计算过程,提炼其中最优的平行论元对,以此构建辅助平行推理机制的线索集合  $U_{key}$ (即图 4 中的线索集合  $U_{key}$ ).



Fig.6 An example of candidate parallel ArgPair selection

图 6 候选平行论元对采集样例

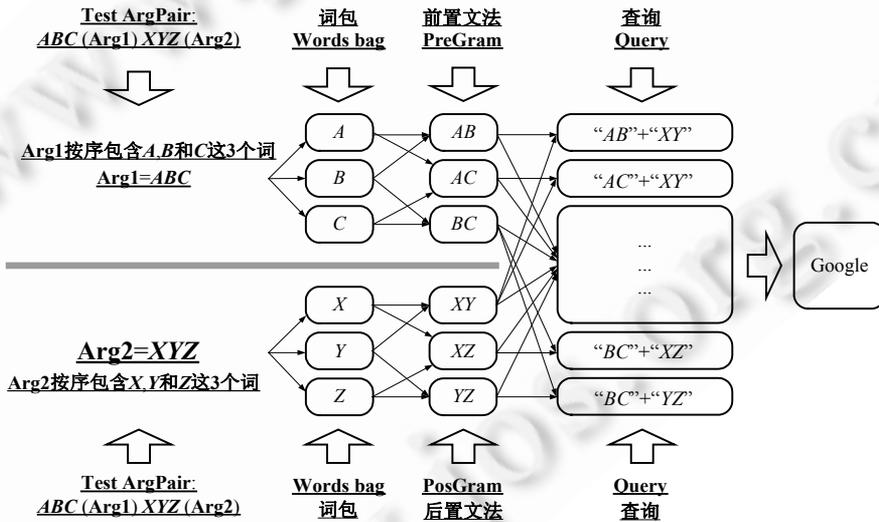


Fig.7 Procedure of building the set of candidate parallel ArgPairs

图 7 候选平行论元对集合构建过程

### 5.2 语义平行度计算

语义平行度计算的目的是:度量上述候选平行论元对与待测论元对的语义一致性强弱.值得注意的是,本文所指的论元是一种反映篇章关系的表现单元,表现为诸如文字块、子句或句子的文字片段,而不是语义角色标注中“谓语-论元”组块中的“论元”(即具有论元角色的词或短语).此处需对 PDTB 的论元称谓和语义角色中的论元称谓区别对待.相应地,本节所涉的语义平行度计算,是一种以语块、子句或句子的整体搭配为单位进行相似度计算的过程,而不是以某一孤立的角色单元进行匹配.下文提到的角色加权方法,是将论元(即语块、子句或句子)中每个词特征的角色作为语义一致性的附加计算因子,而不是“谓语-论元”组块中“论元”的一致性匹配.下面首先阐述语义平行度计算的必要性,然后给出具体的计算方法.语义平行度计算的必要性包括如下两个方面:

- (1) 某些二元文法 PreGram 和 PosGram 无法表征待测论元对的语义,以此形成的查询往往在检索过程中引入不相关反馈,从这类反馈中采集的候选平行论元对实际上与待测论元对具有完全不同的语义,从而成为线索集合  $P$  中的噪声,误导基于“论元语义平行则关系平行”的推理机制。
- (2) 如第 4.1 节所述,两两论元对语义的平行程度不同,则两者关系的平行概率也不同.因此,度量语义平行度有益于估计不同候选与待测论元对的关系平行概率,辅助选取更优的候选及其关系参与平行推理过程。

语义平行度计算包含两个部分:其一是平衡计算模式,其二是平行计算模型.其中,平衡计算模式尝试解决前置论元 Arg1 和后置论元 Arg2 在语义平行度计算过程中的作用平衡问题.如图 8 所示,两个论元对前置论元 Arg1 的平行度为  $PreSim$ ,后置论元 Arg2 的平行度为  $PosSim$ ,经过  $PreSim$  和  $PosSim$  数值上的权衡,形成论元对整体之间的平行度( $Sim$ ),计算方法如下:

$$Sim = \frac{PreSim + PosSim}{2 \cdot e^{|PreSim - PosSim|}} \quad (1)$$

其中,分母的指数运算起归一化作用.此外,无论  $PreSim$  或  $PosSim$  都采用相同的语义平行度计算方法.该方法以统一的向量空间(vector space model,简称 VSM)作为单论元的语义描述结构,借助基于语义角色加权的 Jaccard 算法<sup>[15]</sup>进行平行度计算.Jaccard 算法如公式(2),加权的 Jaccard 算法如公式(3):

$$J = \frac{a_i \cdot a_c}{|a_i| + |a_c| - a_i \cdot a_c} \quad (2)$$

$$J_R = \frac{R(a_i) \cdot R(a_c)}{|R(a_i)| + |R(a_c)| - R(a_i) \cdot R(a_c)} \quad (3)$$

其中, $J$  表示基本 Jaccard 度量值; $a_i$  表示待测论元对中 Arg1 或 Arg2 的特征向量(一元词特征), $a_c$  表示候选论元对中 Arg1 或 Arg2 的特征向量(详解:当计算  $PreSim$  时, $a_i$  为待测论元对中的 Arg1, $a_c$  为候选论元对中的 Arg1;当计算  $PosSim$  时, $a_i$  为待测论元对中的 Arg2, $a_c$  为候选论元对中的 Arg2);分子为内积运算,即  $a_i$  和  $a_c$  对应维度的权值各自求积,然后取和;分母部分的绝对值表示特征向量的长度,即各维度权值自取平方,取和后求平方根;特征向量各维度的权值通过 TFIDF 予以计算,其中的特征频率 TF 为词特征在论元中出现的频率(经停用词过滤后所得),对于待测论元对中的词特征,其频率 TF 依赖该论元对所在篇章(即 PDTB 给出的上下文)进行统计计算,对于候选论元对中的词特征,其频率 TF 则根据该论元对所在网页的正文进行统计计算.反文档频率 IDF 通过大规模静态语料库 GIGAWORD<sup>[16]</sup>(包含 896 446 篇英文新闻文本)进行统计所得. $J_R$  表示角色加权的 Jaccard 度量值,其中的  $R(a_i)$  和  $R(a_c)$  分别表示  $a_i$  和  $a_c$  经过角色加权后的向量空间。

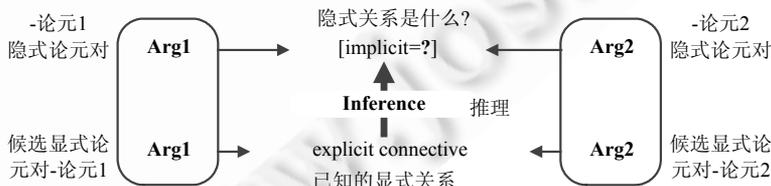


Fig.8 Tradeoff parallelism calculation of ArgPairs

图 8 论元对平行度的平衡计算模式

角色加权的计算过程如下:假设某一论元  $a$  的向量空间为  $T = \{ \langle t_1, w_1 \rangle, \dots, \langle t_n, w_n \rangle \}$ ,将  $a$  进行语义角色标注,生成  $a$  的角色空间  $R = \{ \langle r_1, w'_1 \rangle, \dots, \langle r_n, w'_n \rangle \}$ ,其中, $r_i$  表示特征  $t_i$  的语义角色(施事、受事和谓语等角色).针对向量空间  $T$  的每个特征维度进行角色加权,形成角色加权后的向量空间  $R(a)$ ,计算公式如下:

$$\begin{cases} w'_i = p(r_i | a) \cdot w_i, i \in 1, \dots, n \\ R(a_i) = \{ w'_1, \dots, w'_i, \dots, w'_n \} \end{cases} \quad (4)$$

其中, $w_i$  表示特征  $t_i$  上的原始权重(上述 TFIDF 计算所得); $w'_i$  表示特征  $t_i$  经过角色加权后的新权值; $p(r_i | a)$  是给定

向量空间  $a$  这一先验条件的情况下,特征  $t_i$  具有角色  $r_i$  的概率,换言之,这一概率是:当  $t_i$  与向量空间  $a$  中的其他特征(词特征)共现于同一论元时, $t_i$  在这一论元中充当角色  $r_i$  的概率,其计算公式如下:

$$p(r_i | a_i) = \sum_{\substack{j=1, \dots, n \\ j \neq i}} \log(\arg \max. p(r_i | t_i, t_j) p(t_i | t_j)) \quad (5)$$

其中,  $p(t_i | t_j)$  为特征  $t_i$  与  $t_j$  的共现概率;  $p(r_i | t_i, t_j)$  为特征  $t_i$  与  $t_j$  共现时,  $t_i$  具有角色  $r_i$  的概率. 本文对 GIGAWORD 中 896 446 篇文本的句子进行语义角色标注, 标注过程使用斯坦福句法分析器和 Senna-v3.0 角色标注器(<http://ml.nec-labs.com/senna>)(标注准确率高于 70%), 由此生成的大规模角色标注及词共现样本能够有效地辅助上述角色加权计算.

本文统一采用上述方法对  $a_i$  和  $a_c$  进行角色加权, 为了确保  $R(a_i)$  和  $R(a_c)$  具有统一的角色空间, 本文采用  $a_i$  (即待测论元) 的语义角色标注结果构建角色空间  $R = \{r_1, \dots, r_n\}$ . 从而,  $R(a_i)$  和  $R(a_c)$  的区别主要集中于先验条件  $a$  的向量空间, 详细来说: 在计算  $R(a_i)$  时, 先验条件为  $a_i$ , 加权概率为  $p(r_i | a_i)$ , 即给定待测论元的共现词特征时, 特征  $t_i$  的角色为  $r_i$  的概率; 在计算  $R(a_c)$  时, 先验条件为  $a_c$ , 加权概率为  $p(r_i | a_c)$ , 即给定候选论元的共现词特征时, 特征  $t_i$  的角色为  $r_i$  的概率. 由此, 利用角色加权的  $R(a_i)$  和  $R(a_c)$  进行 Jaccard 计算时, 如公式(3), 所获数值实际上反映了候选论元  $a_c$  中词特征的角色, 趋近于这类特征在待测论元  $a_i$  中的角色的概率.

本文借助上述查询构建和语义平行度计算, 以 Google 搜索引擎为平台, 实现了检索驱动的候选平行论元对的挖掘方法. 利用这一方法, 对每个待测论元对挖掘候选样本, 形成对应的候选集及后续线索集.

## 6 隐式关系的平行推理算法

检索驱动的平行论元对挖掘, 能为分析待测论元对的隐式关系提供参考资源(即候选平行论元对集合). 利用这一资源, 本节给出基本的平行关系推理算法以及优化这一推理的平行概率模型.

### 6.1 平行关系推理

给定某一待测论元对  $X$  (注: 隐式关系的论元对) 以及挖掘所得的候选平行论元对集合  $U$ , 平行关系推理流程分为如下 3 个步骤: 1) 抽取线索集  $U_{key}$ ; 2) 关系映射; 3) 关系推理, 如图 9 所示.

**Input:**  $X$  and  $U$  (输入待测论元对  $X$  和候选集  $U$ ).

**Basic inference** (基本推理):

Step 1: 从  $U$  中择优抽取平行论元对形成线索集  $U_{key}$ ;

Step 2: 利用显式连接词与篇章关系的映射表(见表 2), 获得  $U_{key}$  中所有论元对的显式关系;

Step 3: 统计获得  $U_{key}$  中出现频次最大的关系  $\arg \max. R_{key}$ ;

**Output:**  $\arg \max. R_{key}$  (极大似然关系  $R_{key}$ ).

Fig.9 Basic parallel inference procedure

图 9 基本的平行推理流程

其中, 线索集  $U_{key}$  的抽取过程如下: 首先, 根据上节所述的语义平行度  $Sim$  (如公式(1)~公式(5)), 对候选平行论元对集合中的样本进行排序, 形成语义平行度由大至小的候选平行论元对列表, 选取列表中排序靠前的  $N$  个 (即 TopN) 候选样本形成线索集  $U_{key}$ . 换言之, 选择平行度最高的  $N$  个候选样本作为推理线索. 步骤 2 则利用平行论元对具有显式关系的特点(如图 6 所示), 通过显式连接词与篇章关系的映射表(见表 2), 获取  $U_{key}$  中每个论元对的篇章关系类型. 在此基础上, 步骤 3 检验  $U_{key}$  中出现频次最多的一项篇章关系, 并将这一关系判定为待测论元对  $X$  的篇章关系. 比如, 假设“ $I$  hate you (Arg1)  $I$  love you (Arg2)”为待测论元对, 其 TopN( $N=10$ ) 的推理线索集  $U_{key}$  见表 3, 其中, 最大似然关系为“Comparison”(即, 出现频次最多的关系为“对比”关系), 由此推理待测论元对的隐式关系为“Comparison”.

值得注意的是, 这一过程的步骤 2 使用了连接词与篇章关系的显式映射, 而如第 4.1 节所述, 部分连接词本身具有歧义, 无法在步骤 2 中直接指定论元对的篇章关系, 进而也难以支持步骤 3 的最大似然关系估计. 针对这一问题, 本文约略了  $U_{key}$  中歧义连接词对应的关系线索, 包括“since”, “while”和“as”等 11 项连接词引发的线索样

本.比如,表 3 中的 Top4,Top7 和 Top10 将被约略,不参与最大似然关系估计.约略后产生的空缺,由排序靠后的线索依次上补.

Table 3 Example of cue set  $U_{key}$ 表 3 线索集  $U_{key}$  样例

Top1	I hate you	【but】	I love you	(Comparison)
Top2	I hate you	【then】	I love you	(Comparison)
Top3	I hate you	【but】	I love you more	(Comparison)
Top4	I don't hate you	【while】	I love you	(Comparison)
Top5	I hated you	【but】	now I love you	(Comparison)
Top6	I really hate you	【although】	I loved you	(Comparison)
Top7	I couldn't hate you	【as】	I always love you	(Contingency)
Top8	He say I hate you	【but】	in fact I love you	(Comparison)
Top9	I hate you this time	【and】	I will never love you	(Expansion)
Top10	Have you hated me	【since】	I said I didn't love you	(Temporal)

此外,本文不仅使用了 PDTB 的 100 项连接词(最大似然估计过程中实际使用了 89 项无明显歧义的连接词),同时也通过人工收集与标注,扩充了连接词集合,共包含 183 项无歧义的新词,借以提高有效线索的召回能力.扩充方法如下:(1) 以 PDTB 中已知的连接词为种子,利用 WordNet(<http://wordnet.princeton.edu/wordnet/download>) 的上下位和同义关系获取词义相近的新词.(2) 人工标注其中具有语义连接作用的词,包括连接词和功能性连接词,其中,功能连接词本身不是连接词,但具有语义关联的作用.比如,“result in”不是连接词,但能反映论元的因果关系.在功能连接词的使用过程中,本文利用斯坦福开发的句法依存分析器(stanford parser(<http://nlp.stanford.edu/software/lex-parser.shtml>)),对作为线索的毗连论元对进行依存分析,在确定两论元同时依存于功能连接词时,才判定这一线索为有效项.即,功能连接词反映的关系确为论元之间的关系,而不是单一句子中句法成分的关系.

## 6.2 平行概率模型

如第 4.1 节所述,两两论元对的语义平行程度不同,则两者篇章关系的平行概率也不同.平行概率模型即是针对这一问题提出的处理方法,其目标是将两两论元对的语义平行度引入两者关系的平行概率计算当中,借以优化上述利用极大似然关系的推理方法.

事实上,如果  $N$  值选择合理,则上述线索集  $U_{key}$  的最大似然关系平行于待测关系(即,待测论元对的关系)的概率较高,由此形成的推理结果也较为准确.但是,不同待测论元对的线索集  $U_{key}$  往往具有不同的内部关系分布,从而,当采用同一  $N$  值进行最大似然关系判定时,不均衡的关系分布会导致推理偏见.比如,表 4 给出了两个待测论元对  $X$  和  $Y$  的线索集,并给出两者经过平行度  $Sim$  排序后的 Top10 结果.如果已知  $X$  和  $Y$  的隐式关系都为“Comparison”,可以发现,利用 Top10 中的最大似然关系指定待测关系,只会使得  $X$  获得正确推理,而利用 Top3 中的最大似然关系,则只会使  $Y$  获得正确推理,从而形成了推理偏见.

Table 4 A contrast example of the distribution of relation types between different cue sets

表 4 不同线索集内部关系分布对比样例

Test ArgPair X(测试论元对 X)		Test ArgPair Y(测试论元对 Y)	
Top1: Comparison	( $J_R=0.300$ )	Top1: Comparison	( $J_R=0.300$ )
Top2: Contingency	( $J_R=0.290$ )	Top2: Comparison	( $J_R=0.290$ )
Top3: Contingency	( $J_R=0.280$ )	Top3: Comparison	( $J_R=0.280$ )
Top4: Comparison	( $J_R=0.100$ )	Top4: Expansion	( $J_R=0.100$ )
Top5: Comparison	( $J_R=0.009$ )	Top5: Expansion	( $J_R=0.050$ )
Top6: Comparison	( $J_R=0.008$ )	Top6: Temporal	( $J_R=0.005$ )
Top7: Comparison	( $J_R=0.007$ )	Top7: Expansion	( $J_R=0.004$ )
Top8: Comparison	( $J_R=0.006$ )	Top8: Expansion	( $J_R=0.003$ )
Top9: Comparison	( $J_R=0.005$ )	Top9: Contingency	( $J_R=0.002$ )
Top10: Comparison	( $J_R=0.004$ )	Top10: Expansion	( $J_R=0.001$ )

注:该例只是为方便分析而构造的参考样例,其中,关系分布和  $J_R$  值并非真实样本所获数据.

推理偏见的根本原因是篇章关系本源的分布失衡,本文对 PDTB 语料库(全部样本)中的显式篇章关系进行统计,发现 PDTB 的 4 种关系大类具有明显的分布失衡性,见表 5.而这一失衡性也基本符合大规模数据资源(如 GIGAWORD 和检索结果)中的关系分布特点.从而,分布较广的关系(如 expansion)总是以较高的概率出现于不同测试样本的线索集  $U_{key}$  中,形成噪声,干扰基于最大似然关系的推理过程.

Table 5 Distribution of explicit relation in PDTB

表 5 PDTB 中的显式关系分布

	Expansion	Contingency	Comparison	Temporal	Total
Explicit	6 431 (33.05%)	3 741 (19.23%)	5 590 (28.73%)	3 696 (18.99%)	19 456
Implicit	8 868 (53.47%)	4 261 (25.69%)	2 505 (15.10%)	950 (5.73%)	16 584

由此,本文给出一种关系平行概率模型,该模型将论元对的语义平行度  $Sim$  和关系分布的先验概率融入关系的平行概率计算,方法如下:给定线索集  $U_{key}$  中出现的某一关系  $c$  以及具有这一关系的所有平行论元对,将这些论元对设为集合  $u$ ,利用集合  $u$  中样本的平行度  $Sim$  和所有关系的先验概率分布,计算  $c$  的平行概率,公式如下:

$$P_c = \sum_{i \in u} Sim_i \cdot H_i \quad (6)$$

其中,  $P_c$  表示平行概率,即集合  $u$  对应的关系  $c$  与待测关系相互平行的概率;  $Sim_i$  表示集合  $u$  中第  $i$  个论元对与待测论元对的角色加权平行度;  $H_i$  表示第  $i$  个论元对近邻的关系分布差异熵,公式如下:

$$\begin{cases} H_i = \sum_{j \in rel} D_j \cdot \log(D_j) \\ D_j = d_j \cdot d_{norm(j)} \end{cases} \quad (7)$$

其中,  $D_j$  表示关系  $j$  的局部分布概率差异,这一差异的计算方式如下:假设线索集  $U_{key}$  中的所有论元对已按平行度  $Sim$  进行排序,以固定窗口大小(window)选取第  $i$  个论元对的上下近邻,将近邻论元对所对应的所有显式关系作为以第  $i$  个论元对为中心的局部关系样本集,度量其中关系  $j$  的出现频率,作为  $j$  的局部关系分布  $d_j$ ,同时,利用基于大规模语料统计获得的先验概率分布  $d_{norm(j)}$  进行差值运算并取绝对值,由此获得  $D_j$ ,从而,  $D_j$  能够反映第  $i$  个论元对的近邻中,关系  $j$  的局部分布与先验分布的差异,这一差异越大,则表明关系  $j$  的局部分布越不稳定;在此基础上,  $H_i$  对所有 PDTB 关系的局部分布差异(以论元对  $i$  为中心的局部关系差异)计算熵值,熵值越大,说明所有关系的局部分布越混乱且越不稳定,也就是说,局部分布整体偏离先验分布的现象越明显,同时也说明,平行度  $J_R$  发挥了有效作用,触发了局部区域的关系分布发生变化,从而局部区域内,基于语义平行度  $Sim$  估计关系平行概率的置信度得以提升.

由此,公式(6)中的平行概率实际上融合了两种测度:一种是某一论元对的平行度  $Sim$ ,平行度越高,则该论元对的关系与待测关系相互平行的概率越高;另一种是这一论元对的关系作为推理线索的置信度,通过论元对近邻的关系分布和先验分布的差异计算所得,差异熵越大,则置信度越高.从而,一个论元对的关系平行概率即是上述两种测度的权衡指标.而公式(6)实际上是对线索集  $U_{key}$  中同一关系  $c$  的所有论元对进行上述概率计算,并最终取和,进而使这一关系的样本数量也发挥平行概率度量的作用.由此形成的  $P_c$  即是关系  $c$  平行于待测关系的概率.在此基础上,本文对线索集  $U_{key}$  中出现的每一种关系进行上述平行概率计算,并判定概率最大的关系作为待测论元对的关系.

## 7 基于同指焦点的隐式关系消歧

如第 4.1 节所述,论元一般都存在论述焦点,上下文无关时,论元的焦点存在多样性(即焦点歧义),当不同焦点触发不同关系时,即引入关系歧义,这类歧义往往难于识别,对关系推理(尤其是隐式关系推理)产生误导(800 个标注样本中存在 22%的焦点歧义).针对这一问题,本文提出一种基于同指焦点的歧义关系检测和屏蔽方法,其遵循如下假设:两两语义平行的论元对只有在焦点相同的情况下,两者的关系才能达到事实上的平行;相应地,在上述平行推理过程中,只有线索集  $U_{key}$  中的平行论元对与待测论元对具有相同焦点时(即同指焦点情况下),这类平行论元对及其显式关系才满足作为推理线索的条件.

本节给出基于同指焦点的隐式关系消歧方法,包括焦点识别、同指焦点检测和相应的消歧过程,以此优化上述平行推理过程。

### 7.1 主题驱动的焦点识别方法

焦点识别问题在国内外尚无成型研究,只有宾州树库(Penn tree bank,简称 PTB)给出了否定词的焦点识别任务和相关研究。而本文涉及的焦点更为开放,现有针对否定词的焦点识别方法不具有普适性。事实上,焦点表现为论元对中一组被着重强调的词或组块,往往反映论元及其所属篇章的作者在陈述论点的过程中表现出的核心思想、观点和情感。如果将主题作为核心思想、观点和情感特征的混淆集,则焦点往往与主题有着较强的相关性。

由此,本文采用如下方法判定论元焦点(focus):给定论元及其所在篇章,论元中某一词特征(下文统称论元特征)与篇章某一关键词的相关度为  $rel_j$ ,则论元特征与篇章主题的相关性  $R$  约等于该特征与篇章所有关键词的相关度  $rel_j$  之积;而相关性  $R$  将用于判定论元特征是否为论元焦点的量化指标,通过与当前论元中其他特征的  $R$  值进行横向对比,实现论元焦点的判定。即相关性  $R$  越大,则论元特征作为论元焦点的可能性越高。给定某一论元特征,  $R$  值计算如下述公式(8):

$$R = \prod_j rel_j \quad (8)$$

其中,  $j$  表示篇章关键词的序号,  $rel_j$  表示论元特征与当前篇章第  $j$  个关键词的相关度。此处所提的篇章,为待测论元对所在的篇章(PDTB 文本),本文利用 TFIDF 作为衡量词特征描述篇章主题思想的量化指标,某一词特征的 TFIDF 指标越高,则该词特征表述篇章主题的能力越强,也越关键。本文对待测论元对所在篇章蕴含的所有词项进行 TFIDF 计算(TF 为词项在篇章中的频率, IDF 为基于大规模静态数据 GIGAWORD 获得的反文档频率),并根据 TFIDF 对词项由大到小排序,选取 TOP10(经验值)的词项形成关键词集合,表征篇章主题。给定待测论元的某一特征和待测论元所在篇章的某一关键词,本文采用两者在大规模语料集中“共现并同为关键词”的概率作为两者相关度  $rel_j$  的度量。对这一度量方法的解释是:如果两个词在大量文本中同时作为关键词共现,那么两者的相关度较高。从而,对于上述焦点识别方法的解释是:如果论元特征与篇章关键词在大规模语料中频繁共现,且两者共现时都为关键词,则两者的相关性较高。而在论元所在的篇章中,论元特征与篇章的所有关键词都有较高相关度,则论元特征与篇章主题相关的概率较高。而相关于主题的概率越高,则论元特征作为论元焦点的概率越高。

### 7.2 同指焦点检测模型

利用焦点识别结果,本文提出一种焦点同指的检测模型,用于判定待测论元对与平行论元对之间的焦点一致性检验,扶助平行推理过程中歧义关系的屏蔽。

同指焦点检测过程如下:首先,假定论元对(包括待测论元对和平行论元对)中的每个词都为候选焦点,并且每个词都对应一项焦点概率,概率计算如公式(8)所示;其次,建立待测论元对和平行论元对的焦点概率曲线,曲线建立于二维坐标系,横坐标为词特征序列,纵坐标为词特征对应的焦点概率,所有论元对采用统一特征序列标定横坐标,从而支持不同概率曲线的可比性;最后,匹配待测论元对和平行论元对的焦点概率曲线,匹配所得的偏差越小,则待测论元对和平行论元对具有相同焦点(即同指焦点)的概率越高。这一过程涉及焦点概率曲线的量化描述及其匹配计算。

本文采用一元多阶多项式进行焦点分布曲线的量化描述,使多项式形成的连续曲线尽量逼近二维空间中离散分布的焦点概率,多项式如下所示:

$$f(x) = \sum_{i \in m} a_i \cdot x^i \quad (9)$$

其中,  $m$  表示多项式的最高次幂,  $a_i$  为逼近系数,通过训练差异函数  $C(a)$  予以计算,  $C(a)$  表示特定  $a_i \{i=1, \dots, m\}$  取特定数值后,曲线  $f(x)$  在特征序列上的取值(纵坐标)与实际焦点概率  $R$  值的方差,这一度量计算了焦点分布曲线与真实值的差异:

$$C(a_1, \dots, a_m) = \sum_{i \in n'} [R_i \cdot f(x_i)]^2 \tag{10}$$

其中,当方差取到最小值时, $a_i \{i=1, \dots, m\}$  获得最优解; $n'$  表示  $R$  值非零的样本点数目.为了避免过拟合,本文将  $m$  设置为略小于  $n'$  的数值( $m=n'-1$ ).在此基础上,通过计算  $C(a)$  关于  $x$  的偏导数获得参数  $a_i \{i=1, \dots, m\}$  的数值.焦点概率曲线的匹配则采用差值积分予以计算.给定待测论元对的概率曲线  $f_i(x)$  和平行论元对的概率曲线  $f_r(x)$ ,差值积分计算如下:

$$D = \int_1^n |f_i(x) \cdot f_r(x)| dx \tag{11}$$

其中,积分  $D$  表示两曲线的差值区域,如图 10 中的灰色区域.因此,当待测论元对与平行论元对的  $D$  值(即灰色区域)最小时,则两者具有相同焦点的概率最大,即焦点同指性最高.

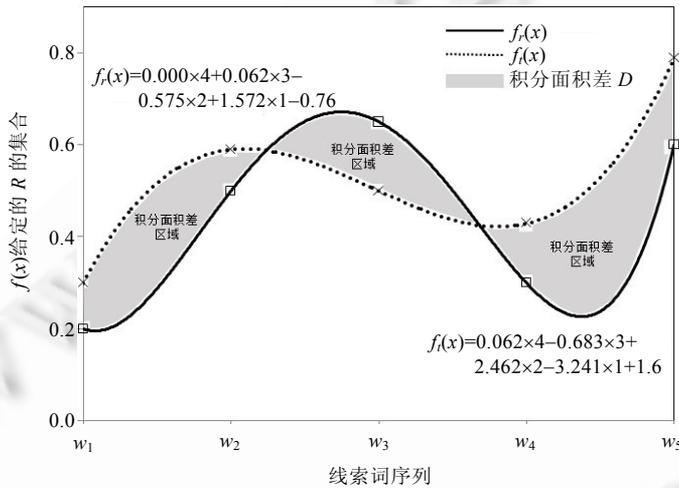


Fig.10 Deviation integration of focus-probability curve  
图 10 焦点概率曲线的差值积分

### 7.3 基于同指焦点的关系消歧

同指焦点概率模型能够测量一个平行论元对与一个待测论元对的焦点概率一致性,值得注意的是:这一概率模型将论元对中的每个词特征都看作一个可能的焦点,并赋予了数值化的焦点概率,从而,短语、多词组块或掺杂停用词的离散词序列也可参与焦点概率计算和匹配,并支持层次性的多焦点匹配.

本文利用这一概率模型对上述线索集  $U_{key}$  中的歧义关系进行消歧,消歧策略是利用同指焦点概率进一步约束公式(6)中的  $P_c$  指标,这一指标本身反映了关系  $c$  的平行概率,即,关系  $c$  平行于待测论元对关系的概率,并包括了语义平行度  $Sim$  和关系置信度  $H_i$  两种测度.本节在此基础上融入同指焦点概率  $D$ ,计算公式如下:

$$P'_c = \sum_{i \in u} sim_i \cdot H_i \cdot D_i^{-1} \tag{12}$$

其中, $u$  为线索集  $U_{key}$  中具有显式关系  $c$  的平行论元对子集; $D_i$  是  $u$  中某一平行论元对与待测论元对的同指焦点概率,公式中该指标取反,以保证与  $Sim$  和  $H$  指标一样,对  $P_c$  值产生一致的大小影响( $Sim$  和  $H$  指标越大,则  $P_c$  指标越优; $D$  值越小,则  $D^{-1}$  值越大, $P_c$  指标越优).从而, $U_{key}$  中出现的某一已知关系  $c$  (即显式关系)是否与待测论元对的隐式关系平行,并作为推理结果指定待测关系,取决于关系  $c$  的 3 种测度:1) 语义平行度;2) 关系置信度;3) 反映焦点一致性的同指焦点概率.利用优化的  $P_c$  指标,可对线索集  $U_{key}$  中出现的所有关系进行概率权衡,并选择  $P_c$  指标最高的关系指定待测论元对的隐式关系,形成最终的推理结果.

## 8 实验设计

本节给出实验使用的语言学资源、评测方法以及参与实验的隐式关系检测系统.

### 8.1 语料

本文采用第2版 PDTB<sup>[2,17,18]</sup>的第21节、第22节为测试语料1(TD1),共包含1 046个样本(英文隐式关系论元对),同时采用第23节、第24节作为测试语料2(TD2),共包含1 192个样本,其余第1节~第20节中的14 346个样本作为训练语料.这一语料的配置方法继承了近期权威研究的惯用设置方法,比如 Zhou 等人(Coling 2010)使用 TD1 进行测试,Wang 等人(ACL 2011)使用 TD2 进行测试,从而,这一设置适于进行对比实验.本文方法对训练语料的使用方法如下:(1) 统计连接词与篇章关系的映射概率,见表2.(2) 模型参数的训练,包括  $n$  元文法的  $n$  值,见第5.1节;TopN 的  $N$  值(出现于线索集  $U_{key}$  的生成过程中),见第6.1节;局部分布差异计算中的近邻窗口大小(window),见第6.2节.

实验涉及的外部资源包括两类:一类是大规模静态数据资源 GIGAWORD<sup>[16]</sup>,包含4 111 240篇来自纽约时报等媒体的新闻报道,本文抽取其中896 446篇英文文本(源于检测任务仅涉及单语英文样本)进行词特征 TFIDF、语义角色和焦点相关性的概率统计,如公式(2)、公式(4)和公式(8),这一语料也用于健壮性分析过程中本地平行论元的检索;另一类外部资源是媒体和类型自由的 Web 文本,借助搜索引擎 Google,扶助挖掘候选平行论元对,见第5.1节.

### 8.2 评测方法

本文使用准确率(accuracy)进行隐式篇章关系检测系统的性能评测,准确率是目前篇章关系分析与检测研究领域的通用评测标准,其计算模式如下:

$$Accuracy = \frac{TruePositive + TrueNegative}{All}$$

其中,Positive 表示系统判定待测论元对具有某一关系,Negative 表示系统判定待测论元对不具有某一关系, True 表示系统判定正确,False 表示系统判定错误(false 在准确率 accuracy 中没有应用).从而,TruePositive 表示系统正确判定待测论元对具有某一关系,TrueNegative 表示系统正确判定待测论元对不具有某一关系,All 则表示待测论元对的总数.评测过程统计其中 TruePositive 和 TrueNegative 的论元对数量,求和后除以 All,形成准确率 Accuracy.

该评测设计之初的目的是计量单一关系的二元分类性能,比如,以表1中的第1层主关系为例,如果单一关系“Expansion”为一类,则另外3种关系(包括“Contingency”,“Temporal”和“Comparison”这3种关系)为另一类,并被视作非“Expansion”关系类.当给定特定论元对时,其或者具有“Expansion”关系或具有非“Expansion”关系,那么,如果论元对具有“Expansion”关系且系统判定其具有“Expansion”关系,则获得 TruePositive 指标,反之,如果系统判定为“Contingency”,“Temporal”或“Comparison”这3种关系中的任何一种,都为误判,即 FalseNegative;相对而言,如果该论元对具有非“Expansion”关系,则无论系统判定为“Contingency”,“Temporal”或“Comparison”中的哪一种,只要不是“Expansion”,则都为判定正确,并获得 TrueNegative 指标,否则为误判,即 FalsePositive.

目前,这一准确率也用于多元关系的性能计量,计量过程将 TrueNegative 设置为恒定0值,只检验每个论元对是否获得正确的关系判定,即,只计算 TruePositive 指标与 All 的比值,并将其作为准确率.本文实验部分只选择 PDTB 关系体系中的第1层关系进行评测,即“Expansion”,“Contingency”,“Temporal”和“Comparison”这4类主关系(见表1),由此,上述多元关系评测即为四元关系评测.下文遵循领域惯用的表示方法,将这种四元主关系的准确率标定为“four-way”评测指标.

本文仅选择待测论元对的第1层关系作为评测对象,原因在于:现有隐式关系检测方法在第1层上仅能获得约40%的准确率,而第2层以后的关系检测在很大程度上依赖第1层检测的性能优劣,从而在有效提升第1层检测精度之前,评测后续层级的关系并无实际意义.

### 8.3 实验系统

本文面向隐式篇章关系检测任务,提出了检索驱动的平行推理机制以及优化这一机制的语义平行度计算方法、关系平行度计算方法和关系消歧方法.由此,实验部分在测试平行推理效果的同时,对上述优化方法的性能也分别进行了评测,并将整体性能与现有方法进行对比测试.相应地,参与实验的系统如下:

- 基准系统(baseline)
  - RB 系统:随机检测(random baseline),给定测试样本集,随机指定每个待测论元对的关系,以 {Expansion,Contingency,Temporal,Comparison} 为关系序列,利用随机函数抽选关系.
  - MINB 系统:最小概率检测(minimum baseline),给定测试样本,判定所有待测论元对都具有唯一一种关系,且为出现概率最小的关系,概率最小关系通过训练语料获得为“Temporal”(见表 5).
  - MAXB 系统:最大概率检测(maximum baseline),给定测试样本,判定所有待测论元对都具有一种关系,且为出现概率最大的关系,概率最大关系通过训练语料获得为“Expansion”(见表 5).
- 平行推理系统
  - PBI 系统:基本平行推理(basic parallelism inference),给定测试样本集,针对每个待测论元对,抽取  $n$  元文法( $n=2$ )形成查询,利用搜索引擎获取候选平行论元对,利用基于 Jaccard 算法的平衡相似度  $Sim$ (如公式(1)、公式(2))对候选进行排序,选取 Top $N$  个候选平行论元对建立线索集  $U_{key}$ ,利用连接词与显式关系的映射表(见表 2)获取  $U_{key}$  中所有论元对的显式关系,并将其中出现最频繁的关系作为待测论元对的最大似然关系,形成显式指导隐式的平行推理.
  - JR+BPI 系统:角色加权的语义平行度优化( $J_R$  based BPI),利用角色加权的 Jaccard 算法优化 BPI 系统的候选平行论元对排序.换言之, JR+BPI 系统继承了 BPI 系统的基本流程,只在利用 Top $N$  形成线索集  $U_{key}$  之前对每个候选平行论元对的平衡相似度  $Sim$  进行优化,使用了角色加权后的  $J_R$  予以计算,如公式(3)~公式(5).
  - PC+BPI 系统:关系平行概率优化( $P_c$  based BPI),该系统也继承了 BPI 系统的基本流程,区别在于,该系统没有采用线索集  $U_{key}$  中的高频关系作为待测论元对的关系,而是对  $U_{key}$  中出现的每种关系进行平行概率  $P_c$  的计算,如公式(6)、公式(7),并采用概率较高的关系作为待测论元对的关系.
  - FD+BPI 系统:同指焦点概率优化(focus-deviation based BPI),该系统在 BPI 系统中融入了关系消歧模块,消歧过程主要施加于关系平行概率计算过程.即,采用同指焦点概率约束关系平行概率,形成消歧后的平行概率  $P'_c$  如公式(9)~公式(12).

上述 PC+BPI 和 FD+BPI 系统都未采用  $U_{key}$  中的高频关系进行平行推理,而是分别使用了全新的关系平行概率模型.尽管如此,两者却都需要线索集  $U_{key}$  进行后续概率计算.而线索集  $U_{key}$  的优劣受制于语义平行度计算的质量,即受制于 BPI 系统中的 Jaccard 算法.为此,实验额外地建立了 JR+PC+BPI 和 JR+FD+BPI 系统,旨在将语义平行度优化的效果融入 PC+BPI 和 FD+BPI 系统.换言之, PC+BPI 和 FD+BPI 系统使用的线索集  $U_{key}$ ,是利用基本 Jaccard 算法度量的语义平行度;而 JR+PC+BPI 和 JR+FD+BPI 系统使用的线索集,则是使用角色加权的 Jaccard 算法(即  $J_R$ )进行语义平行度计算.关系推理系统的配置情况见表 6.

Table 6 Experiment system configuration

表 6 实验系统配置

System	$J$ (Jaccard)	$J_R$ (角色加权 $J$ )	$Sim$ (平衡计算)	$H(\rightarrow P_c)$ (局部差异熵)	$D(\rightarrow P'_c)$ (焦点差值积分)	$U_{key}$
BPI	Y	-	Y	-	-	Y
PC+BPI	Y	-	Y	Y	-	Y
FD+BPI	Y	-	Y	Y	Y	Y
JR+BPI	-	Y	Y	-	-	Y
JR+PC+BPI	-	Y	Y	Y	-	Y
JR+FD+BPI	-	Y	Y	Y	Y	Y

• 对比系统

本文选择 Wang 等人<sup>[4]</sup>和 Zhou 等人<sup>[5]</sup>的隐式篇章关系检测方法进行对比实验,前者主要利用语言学特征实现关系分类,后者采用统计策略实现关系推理,基本代表了篇章分析领域的主流研究思路和方法论,具有性能对比的实际价值,系统描述如下:

- ParserSVM 系统:该系统由 Wang 等人提出,给定一项论元对,该系统采用 SVM(支持向量机)对其进行关系分类,这一系统强调句法结构特征的分类作用,并使用树核函数对句法特征进行了优化.
- PatternINF 系统:该系统由 Zhou 等人提出,该系统采用三元文法(trigram)建立关系模板,利用语言模型在 TDT 语料库(topic detection and tracking corpus)上训练关系模板,测试时,衡量待测论元对与模板的生成概率,并利用这一概率选取关系进行推理.

注:系统名称 ParserSVM 和 PatternINF 不是原作的命名,原作无命名,此处命名仅为系统区分之用.

## 9 实验结果及分析

本节首先汇报上述平行推理系统的性能,并给出相应的结果分析;其次,通过与现有方法的对比分析,探讨平行推理机制的优缺点;最后,对检索驱动平行推理系统进行健壮性分析.

### 9.1 平行推理性能及分析

• 训练结果

实验通过训练,获得如下参数取值:(1)  $n$  值为 2,即,使用二元文法进行查询构造,如第 5.1 节;(2)  $N$  值为 20(+9),即,从候选平行论元对集合  $U$  中抽取  $Sim$  值(如公式(1))最高的 29 个样本,加入线索集合  $U_{key}$ ,但其中实际参与推理测试的样本仅为前 20 项,后 9 项不作为推理线索,而是作为附加项,防止局部分布差异熵  $D_j$  的窗口(window)为空,见第 6.2 节;(3) 窗口  $window$  取值为 10,即局部分布差异熵  $D_j$  的计算过程中,每个平行论元对的近邻数设置为 9(包括其自身则共计为 10 项).由于  $U_{key}$  中参与推理的平行论元对共计 20 项,且近邻选取须向后平移 9 个样本,因此在设置  $U_{key}$  大小时,需附加 9 项近邻(达到 29 项样本),借以防止排序靠后的平行论元对无法从窗口中获得足额的近邻.

此外,针对每个待测论元对蕴含的所有查询(即二元文法组合 PreGram+PosGram),统一将检索结果中的前 500 项作为挖掘候选平行论元对的数据源(如第 5.1 节的挖掘规则).经统计发现:当采用 500 项检索结果进行候选挖掘时,绝大部分查询能够给出均衡的候选数量,维持在 15~25 之间.如果将这一数值设置过小,则存在大量查询无法给出有效候选论元对的现象(有效候选论元对即为满足第 5.1 节挖掘规则的论元对);如果设置过大,则引入大量噪声.

训练中使用的系统、参数取值范围及训练结果见表 7.

**Table 7** Test results (test corpus: TD1, parallelism inference systems, Google search in Jan., 2013)

表 7 测试性能(测试语料为 TD1,系列平行推理系统,Google 检索于 2013 年 1 月)

Parameter	System	$U$	$U_{key}$	Training body				
				$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
$n$	BPI	Y	-	0.349 5	<b>0.433 0</b>	0.151 5	0.063 0	0.059 5
$N$	BPI	-	Y	$N=5$ 0.359 5	$N=10$ 0.412 5	$N=20$ <b>0.530 0</b>	$N=50$ 0.399 5	$N=100$ 0.406 0
Window	PC+BPI	-	Y	$window=3$ 0.434 5	$window=5$ 0.501 5	$window=10$ <b>0.556 5</b>	$window=15$ 0.515 0	$window=20$ 0.515 0

训练使用了 2 000 个论元对样本,其中,训练  $n$  值的过程仅使用最基本的推理系统 BPI,且没有使用线索集  $U_{key}$ ,而是使用未加任何筛选的候选平行论元对集合  $U$ .其原因在于, $U$ 中为纯粹的检索结果,有利于检验哪一种  $n$  元文法构造的查询,更有利于借助搜索引擎捕获更多的平行论元对.如表 7 所示,当  $n$  值为 2 时,利用集合  $U$  进行关系推理的 BPI 取得了最优性能,说明此时的检索过程能够获得较多的平行论元对进入  $U$  集,提高了最大似然关系估计的精度,相较而言,当  $n$  值过小时,对语义平行度要求过低,从而检索结果中非语义平行的论元对随之增

多,相应地,基于“语义平行则关系平行”假设的推理将受到干扰;反之,当  $n$  值过大时,虽然语义平行度得以提高,但字面匹配的要求也更为严格,匹配所得的检索结果极为稀疏,甚至绝大多数查询难以获得检索结果( $U$  集为空),无法进行关系推理。

参数  $N$  的训练过程仍然使用系统 BPI,区别是其采用上述训练所得的最优文法(即二元文法, $n=2$ )进行查询构造与检索,并使用了线索集  $U_{key}$ ,其目的是检验  $U$  中多少个排序靠前的平行论元对(TopN)能够给出较佳的最大似然关系推理。如表 7 所示,当  $N$  值为 20 时,推理性能最优。若  $N$  值过小,则关系推理的随机性增加,从而性能较差;反之,若  $N$  值过大,则将向  $U_{key}$  中引入过多噪声,这类噪声多数为字面上与查询(即 PreGram+PosGram)匹配,而语义上与待测论元对毫无关系的检索结果,当平行论元对数量极为有限时,检索系统为了提高召回率而大量反馈的信息中,不平行的论元对数量远多于平行论元对,从而, $N$  值越大,选入  $U_{key}$  的噪声也随之增多。当这类噪声持续增多并达到一定量级时, $U_{key}$  中的关系分布也逐渐趋近于大规模语料中的本源分布,而其中本源分布最广的关系,将始终作为最大似然关系,并由此成为平行推理结果,形成趋近不变的推理性能。比如,训练语料中“Expansion”关系的分布约为 40%左右,而表 7 中,当  $N$  扩大至 50~100 时,BPI 的性能也都趋近于 40%。其原因即在于大规模语料中,“Expansion”关系本源地分布最广(见表 5),当进入  $U_{key}$  的噪声过多时( $N$  值过大所致), $U_{key}$  内的关系分布也趋于本源化,从而,最大似然关系即为分布最广的“Expansion”关系,所有基于最大似然的推理始终将待测论元对推理为“Expansion”关系,因此准确率也就趋近于训练语料中“Expansion”的概率分布,即 40%。

最后,窗口  $window$  的训练过程使用了系统 PC+BPI,该系统在 BPI 的基础上融入了局部分布差异熵,而  $window$  是差异熵计算时使用的近邻数量,从而该系统的性能优劣可直观反映  $window$  的训练效果。这一训练过程中,系统 PC+BPI 已先验地使用了最优的  $n$  值( $n=2$ )和 TopN 值( $N=20$ ),并使用基于线索集  $U_{key}$  的关系平行概率(如公式(6)、公式(7))进行推理。训练发现:当  $window$  的数值为 10 时,PC+BPI 系统的性能最优。相对来说,当  $window$  过小时,特定平行论元对的近邻数量相应过少,这种依赖小样本获得的局部分布差异熵  $H_i$  过于随机,不仅无法优化关系概率计算,反而造成负面影响;当  $window$  过大时(比如  $window=20=N$ ,见表 7), $U_{key}$  内绝大多数平行论元对都具有近似相同的近邻,从而,各自的局部分布差异熵  $H_i$  都接近  $U_{key}$  全局的分布差异熵,不具有区分性,无法优化各个平行论元对的关系概率。

#### • 测试结果

基于上述参数指标(即  $n=2, N=20, window=10$ ),对实验预设的平行关系推理系统进行评测,获得如表 8 和表 9 的测试结果。其中,表 8 的测试结果使用测试语料 TD1,表 9 的测试结果使用测试语料 TD2,两套测试结果可分别与 Wang 等人和 Zhou 等人的关系检测系统(ParserSVM 和 PatternINF)进行对比,如第 9.2 节。此外,各系统的检索时间皆为 2013 年 1 月(搜索引擎具有时新性,同期检索的数据偏差不大)。

**Table 8** Test results on corpus TD1 (Parallelism inference systems, Google search in Jan., 2013)

**表 8** TD1 为测试语料的测试性能(系列平行推理系统,Google 检索于 2013 年 1 月)

System	Exp./Others	Con./Others	Com./Others	Tem./Others	Four-Way
BPI	0.667 3	0.771 5	0.761 0	0.853 7	0.526 8
PC+BPI	0.701 7	0.775 3	0.782 0	0.859 5	0.540 2
FD+BPI	0.748 6	0.793 5	0.800 2	0.885 3	0.565 9
JR+BPI	0.663 4	0.776 3	0.768 6	0.855 6	0.533 4
JR+PC+BPI	0.685 5	0.778 2	0.773 4	0.870 0	0.553 5
<b>JR+FD+BPI</b>	<b>0.707 5</b>	<b>0.777 2</b>	<b>0.776 3</b>	<b>0.874 8</b>	<b>0.572 6</b>

**Table 9** Test results on corpus TD2 (Parallelism inference systems, Google search in Jan., 2013)

**表 9** TD2 为测试语料的测试性能(系列平行推理系统,Google 检索于 2013 年 1 月)

System	Exp./Others	Con./Others	Com./Others	Tem./Others	Four-Way
BPI	0.625 8	0.708 9	0.742 5	0.882 6	0.480 0
PC+BPI	0.656 0	0.717 3	0.749 1	0.901 8	0.511 7
FD+BPI	0.666 9	0.719 8	0.757 6	0.899 3	0.521 8
JR+BPI	0.645 1	0.720 6	0.750 8	0.873 3	0.493 3
JR+PC+BPI	0.676 2	0.734 9	0.776 0	0.878 4	0.530 2
<b>JR+FD+BPI</b>	<b>0.693 8</b>	<b>0.742 4</b>	<b>0.787 8</b>	<b>0.869 9</b>	<b>0.545 3</b>

表 8 和表 9 都包含了两种测试结果:其一为单关系的二元分类评测,比如,Exp./Other 表示“扩展”关系(expansion)和其他关系(other 包括:contingency,简称 Con.;comparison,简称 Com.;temporal,简称 Tem.)的二元分类;其二为 4 种主关系的多元分类评测(four-way),评测方法见第 8.2 节。

测试结果显示:基于检索驱动的基本平行推理系统 BPI,在 TD1 和 TD2 上分获 0.526 8 和 0.480 0 的多元关系分类精度(four-way).在此基础上,融入平行关系概率模型的推理系统 PC+BPI,在 TD1 和 TD2 上分别获得 1.34%(four-way)和 3.17%(four-way)的性能提高,验证了语义平行度排序和局部分布差异熵的联合作用.其次,借助同指焦点概率,在平行关系概率模型中融入关系消歧,即 FD+BPI 系统,也分别在 TD1 和 TD2 上获得 2.57%(four-way)和 1.01%(four-way)的性能优化,验证了焦点同指的平行论元对在屏蔽歧义关系时的积极作用.最后,借助角色加权的 Jaccard 模型( $J_R$ ),上述 BPI,PC+BPI 和 FD+BPI 系统在 TD1 和 TD2 都获得了相应优化,比如, $J_R$  在 TD1 上将 BPI 系统的准确率(four-way)提高了 0.66%,在 TD2 上提高了 1.33%.从而证明:语义角色有助于度量论元对的语义平行度,并进而积极影响“语义平行则关系平行”的推理过程。

• 测试结果分析

上述测试结果显示了两种主要异常:1) 针对 4 种主关系的二元分类性能差异明显,比如,BPI 系统对扩展关系与非扩展关系(即 Exp./Other)的二元分类精度为 0.6673,而这一系统对时序关系与非时序关系(即 Exp./Other)的二元分类精度高达 0.853 7,相差 18.64%;2) 本文前期工作对 TD1 中 800 个论元对进行了焦点歧义的标注,发现 22%的数据存在焦点歧义引起的关系歧义现象,但利用同指焦点概率进行关系消歧的系统 FD+BPI 及 JR+FD+BPI 却仅仅给出了 1.57%和 1.91%的性能优化,差距较为明显.下面对这两种异常给出分析:

1) 二元分类异常

这一异常的主要成因是测试语料中关系样本分布不均,见表 10,较为明显的分布差异是“Expansion”关系(简称 Exp.)和“Temporal”关系(简称 Tem.),前者的出现概率接近 51%,后者的出现概率仅约 3.4%.测试样本关系的先验分布不均,本源地导致二元分类评测有失公允。

Table 10 Relation distribution in test corpus (TD1 & TD2)

表 10 测试语料中的关系分布(TD1 & TD2)

	Exp.	Con.	Com.	Tem.	Total
TD1(隐式)	531 (50.80%)	297 (28.40%)	182 (17.40%)	36 (3.40%)	1 046
TD2(隐式)	606 (50.84%)	338 (28.36%)	208 (17.45%)	40 (3.35%)	1 192

以 TD1 为例,具有 Tem.关系的论元对仅有 36 项,形成 Tem.关系类,具有其他 3 种关系(包括 Exp.,Con.和 Com.)的 1 010 项论元对都归为非 Tem.关系类.按照二元分类评测的要求,只要非 Tem.关系类的样本不被判定为 Tem.关系类,即使关系推理已经错误(即 Exp.误判为 Con.或 Com.;Con.误判为 Exp.或 Com.;Com.误判为 Exp.或 Con.),仍然作为分类正确的指标参与评测.然而,Tem.关系本源地分布稀疏,大规模数据中的分布仅为 18.99%(显式)和 5.73%(隐式),见表 5,远低于 Exp.的 33.05%(显式)和 53.47%(隐式)以及 Com.的 28.73%(显式)和 15.10%(隐式).从而,无论显式或隐式关系,利用其先验概率分布进行的机器学习(如 Wang 等人<sup>[4]</sup>提出的方法)或推理机制(如本文方法)都将倾向于判定 Exp.,Com.和 Con.这类高概率关系,而不是 Tem.关系.由此,检测系统输出的关系中,绝大部分都为非 Tem.关系,数量极为接近测试语料 TD1 中非 Tem.关系的总和,如表 11(以 BPI 系统输出为例),使得非 Tem.关系的判定精度极高,从而异常提升了 Tem.关系二元分类的精度。

Table 11 Results of binary classification for temporal (viz., Tem.) relation

表 11 “Tem.”关系二元分类结果

Test corpus	NOT Tem. (total=1 010)		Tem. (total=36)	
BPI 输出	Exp.(415)+Con.(197)+Com.(260)	Tem.(118)	Exp.(8)+Con.(3)+Com.(4)	Tem.(21)
	NOT Tem. (total=872)	Tem.(total=36)	NOT Tem.(total=15)	Tem.(total=21)

相比之下,无论在 TD1,TD2 还是大规模外部资源中,Exp.关系的分布始终都偏差不大,从而,Exp.关系类与非 Exp.关系类中正确判定的指标不会过度失衡,避免了类似于 Tem.关系的二元分类异常,见表 12(以 BPI 系统输出

为例).

**Table 12** Results of binary classification for expansion (viz., Exp.) relation

表 12 “Exp.”关系二元分类结果

Test corpus	NOT Exp. (total=515)		Exp. (total=531)	
BPI 输出	Con.(144)+Com.(161)+Tem.(90)	Exp.(120)	Con.(56)+Com.(103)+Tem.(69)	Exp.(303)
	<b>NOT Exp.(total=395)</b>	Exp.(total=120)	NOT Exp.(total=228)	<b>Exp.(total=303)</b>

## 2) 关系消歧异常

关系消歧异常的主要成因包括两项:其一是部分具有歧义关系的测试样本,在进行消歧处理前已获得推理系统的正确判定;其二是焦点识别的准确率不足,影响了基于同指焦点概率的关系消歧。

本文前期工作对 TD1 中 800 个待测论元对(约占 TD1 的 76.5%)进行了焦点和歧义关系的人工标注,参与标注的人员为 3 位从事篇章关系研究的硕士研究生(对该领域的研究时长大于 8 个月),标注过程采用的交叉验证方法,具体步骤如图 11 所示。

**Input:**标注样本(800 项 TD1 中的隐式关系论元对)。

**Blind 1(盲检条件 1):**预先不告知标注目的,即焦点与关系歧义的内在联系为未知条件;

**Blind 2(盲检条件 2):**不给出标注样本蕴含的关系;

**Blind 3(盲检条件 3):**标注样本的论述焦点未知;

**Procedure(流程):**

Step 1:标注者分头标注样本的焦点,不对样本中焦点的个数进行约束(样本上下文禁用);

Step 2:对每个样本,将标注所得的所有焦点取交集,并对游离于交集之外的焦点,采用投票方式重新认定(样本上下文可用),将认定出的焦点加入交集,形成每个样本的焦点集合;

Step 3:标注者重新浏览标注样本,并判断样本是否具有歧义关系,仍然采用 Step 2 中的交叉验证方法,给出具有歧义关系的论元对集合;

Step 4:强制要求标注者对具有歧义关系的样本给出唯一一项确定的关系类型,并利用 PDTB 本源标注的标准关系类型,对每个标注者的标注精度进行评测。

**Output:**输出焦点数据、歧义关系样本和标注精度。

Fig.11 Procedure of labeling focuses and ambiguous relations

图 11 焦点及歧义关系标注流程

标注结果显示:

- 在样本上下文未知的情况下(如图 11 中 Step 1 所示),约 56%的标注样本存在焦点争议;在给定上下文的情况下(如图 11 中 Step 2 所示),焦点争议的样本数缩减至 40%。这一现象说明,篇章主题有助于识别焦点,从而在一定程度上支持了主题驱动的同指焦点识别方法(见第 7.1 节)。
- 此外,经过交叉验证和投票认定(如图 11 中 Step 3 所示),具有焦点争议的样本中,约有 55%的样本存在关系歧义,约占标注样本全集(800 个 TD1 论元对)的 22%,从而可粗略估计 TD1 测试集(共 1 046 个论元对)中,焦点歧义导致的关系歧义样本数应多于 16.9%,由 $(800 \times 22\%) / 1046$  计算所得;
- 最后,借助焦点标注的最终结果,可对标注者的焦点识别进行评测(如图 11 中 Step 4 所示),同时,可利用 PDTB 标准关系类型对标注者的篇章关系检测能力予以评测,见表 13。评测结果显示:焦点识别能力强的标注者,同时也具有较高的关系推理能力(关系判定精度较高),这一结果也支持了基于同指焦点的关系消歧方法(如第 7.3 节)。

从而,测试结果中的关系消歧作用应更为明显,而不是只给予关系推理系统仅仅 1.5%~2%的性能提高。换句话说,焦点歧义导致的关系歧义样本在 TD1 中多于 16%,而本文基于同指焦点的关系消歧却仅仅提高 PC+BPI 系统性能大约 1.5%~2%(远低于 16%),形成了异常的测试结果。

这一异常的原因之一是,部分歧义关系的样本在接受消歧之前已获得了推理系统的正确判定。以 PC+BPI 系统为例,标注样本中的 176 项歧义关系,已有 86 项获得 PC+BPI 系统的正确推理,余下的 90 项约占 TD1 的 8.6%。另一个原因是,本文焦点识别方法的性能尚不完善,其识别精度为 60.04%,在此基础上的关系消歧精度为 52.19%,见表 13。从而,针对上述 90 项剩余的歧义关系样本,FD+BPI 系统(含消歧功能)并不能给予显著的消歧,

仅成功消歧 47 项(提高精度 4.5%),并负面影响了少量(30 项)PC+BPI 系统能够正确推理的样本(降低精度约 2.9%)。尽管如此,观察发现,人工标注的性能也并不理想,焦点识别的最高精度为 78.55%,关系消歧的最高精度为 71.03%,见表 13。因此,焦点识别和关系消歧本身即是具有挑战的研究任务,其优化也将有助于篇章关系的分析与处理,我们未来的工作即是在这一起点上继续开展研究。

**Table 13** Performance of focus identification and relation disambiguation (800 samples in TD1) (%)

**表 13** 焦点识别及关系消歧性能(TD1,800 个标注样本) (%)

	Focus(焦点识别)	Relation(关系推理)
Annotator 1	78.55	71.03
Annotator 2	71.76	63.21
Annotator 3	69.66	62.25
<b>Machine</b>	<b>60.04</b>	<b>52.19</b>

## 9.2 对比实验及分析

本节给出对比实验结果,对比的对象包括 3 个基准系统 RANB,MINB 和 MAXB 以及近期相关研究中有代表性的系统 ParserSVM<sup>[4]</sup>和 PatternINF<sup>[5]</sup>。其中,RANB 随机判定待测论元对的关系,MINB 采用分布概率最小的关系(源于训练语料)指定所有待测论元对的关系,MAXB 采用分布概率最大的关系(源于训练语料)指定所有待测论元对的关系,ParserSVM 是基于句法等特征的关系分类器(使用了树核技术),PatternINF 使用基于语言模型的关系模板进行推理。ParserSVM 倾向于语言学特征的应用,PatternINF 倾向于统计特征的应用,基本代表了现有篇章关系分析与检测的主流方法,适于作为对比对象。

参与对比的平行推理系统包括 JR+BPI, JR+PC+BPI 和 JR+FD+BPI。此外,Wang 等人公布的 ParserSVM 性能源于 TD1-section 23 上的测试,本文平行推理系统则使用 TD1 中全部样本进行测试(包括 section 23 和 section 24);Zhou 等人公布的 PatternINF 性能源于 TD2,相应地,本文平行推理系统也使用 TD2 进行了测试。对比实验结果见表 14,各个平行推理系统在 TD1 上的性能明显优于 ParserSVM,其中, JR+FD+BPI 的性能优势达到 17.26%。此外,平行推理系统在 TD2 上的性能明显优于 PatternINF,其中, JR+FD+BPI 的性能优势达到 13.18%。

**Table 14** Results of comparing experiments

**表 14** 对比实验结果

System	Four-Way (on TD1)	Four-Way (on TD2)
RANB (baseline 1)	0.253 3	0.264 0
MINB (baseline 2)	0.034 0	0.033 5
MAXB (baseline 3)	0.508 0	0.508 4
JR+BPI (ours)	0.533 4	0.493 3
JR+PC+BPI (ours)	0.553 5	0.532 0
JR+FD+BPI (ours)	0.572 6	0.545 3
ParserSVM (Wang)	0.400 0 (only section 23)	-
PatternINF (Zhou)	-	0.413 5

此外,平行推理系统在 TD1 和 TD2 上都优于基准系统,而 ParserSVM 和 PatternINF 则低于基准系统 MAXB。换言之,如果已知分布概率最大的关系(即 expansion 关系),则无需任何分析与处理,只需将所有待测样本先验地指定为这一关系,即可获得超越 ParserSVM 和 PatternINF 的性能。造成这一现象的主要原因是关系样本的分布失衡性,详细来说,测试语料中 Expansion 关系的样本数占到全部测试样本的 50.8%,在已知这一分布的条件下,如果将所有测试样本全部刚性地推断为 Expansion 关系,则 50.8%的测试样本都能获得正确的关系判定,产生 0.508 0 的正确率。虽然这一精度远高于 ParserSVM 和 PatternINF 的性能,但实际上,MAXB 对于其他关系的判定精度皆为 0,是一种因为迎合了“偏见”分布而获得的“欺骗性”精度。然而,这种分布失衡并非个例,除了本文采用的标准测试集,PDTB 的训练样本也存在相同情况,如第 6.2 节表 5 所示的隐式篇章关系中,Expansion 关系的分布高达 53.47%,使得简单而刚性的 MAXB 总能获得较高性能。也因此,本文将 MAXB 所能达到的性能作为一种机器智能的基本底线,超过这一基准的关系检测方法可被认为在关系判定的过程中具有了实际的智能。在测试结果中,ParserSVM 和 PatternINF 的性能低于这一基准,但两者优于 RANB(即随机推理)系统,这说明 ParserSVM

和 PatternINF 系统的确具有关系检测的能力,但还不足以达到智能性。

ParserSVM 系统的性能劣势源于其基础理论的错误。ParserSVM 系统利用句法结构作为篇章关系理解的特征,其基本理论是篇章关系是一种结构化关系,尤其论元对往往是毗连的子句,因此句法结构特征应该能够有效地反映篇章关系类别。但事实上,PDTB 给出的篇章关系是一种语义层面的关系,比如因果关系,而句法结构特征并不能直接反映语义及其关系,从而也无法有效区分不同语义关系的类型。比如,利用最小单位的语义(即词义)分析 Attack 和 Wounded 的关系,可较为容易地判定两者之间的因果关系,而两者的句法结构特征为 VP 和 ADJ,仅仅依赖这一特征标签,即使人为判断,也很难给出恰当的关系判断。换言之,句法结构特征本身并不能有效地反映语义(语义角色是句法标注上层的进一步处理),无语义即无语义关系。因此,句法结构特征可用于辅助分析篇章关系的手段,而不能直接用于关系类型的分类过程。

PatternINF 系统的性能劣势源于静态语言学资源的模板稀疏性,PatternINF 主要利用统计概率和语用学特征,借助语言模型,训练和使用特定篇章关系的文法模板以及模板与篇章关系的映射概率。其中,每个模板都是由一对三元文法构成。在测试时,待测论元对与模板一一匹配,并利用匹配度高的模板估计篇章关系。然而,静态语言学资源的信息量有限,获得的模板并不全面,从而,给定特定测试样例时,模板集合往往本身不具有对称的模板,从而出现无模板可用或无关模板“凑合着用”的局面,必然负面影响关系推理。

相对于检测精度的劣势,ParserSVM 和 PatternINF 系统在检测效率上有较高的优势。

表 15 罗列了 ParserSVM,PatternINF 和本文平行推理系统的计算复杂度(包括训练复杂度和测试复杂度)。如果以词或文法的一次计算为一次复杂系数的累加,那么 ParserSVM 的训练和测试复杂度最低,都为  $O(n)$ ,换言之,系统效率最高。其原因在于,ParserSVM 系统的训练和测试都只需对每个词进行 1 次句法标注和 1 次生成概率的计算,因此,如果语料集包含  $n$  个词,则主要计算量为  $2 \times n$  次,从而复杂级别仅约为  $O(n)$ 。相对地,PatternINF 系统在训练过程中需要对每一种 PreGram(前置文法)和 PosGram(后置文法)进行组合,形成候选模板,并根据实际语料集中样本的文法特征逐一进行模板生成概率的计算,从而,如果训练语料包含  $n$  个论元对,则模板组合数量为  $(n/2) \times (n/2)$  个,即  $n^2/4$  个,使得训练过程的复杂度达到  $O(n^2)$ ;而 PatternINF 系统的测试过程只需使用语言模型对每个测试样本进行一次生成概率的计算,因此复杂度为  $O(n)$ 。

Table 15 Comparing complexities of systems

表 15 系统复杂度对比

Systems	Training time	Test time
ParserSVM (Wang)	$O(n)$	$O(n)$
PatternINF (Zhou)	$O(n^2)$	$O(n)$
JR+FD+BPI (ours)	$O(n^2)$	$O(n^2)$

与 ParserSVM 和 PatternINF 系统相比,本文平行推理系统的计算复杂度较高。其原因在于,无论是训练过程还是测试过程,都需要对每一种 PreGram(前置文法)和 PosGram(后置文法)进行组合,形成查询并进行检索,由此,训练和测试中的复杂度都为  $O(n^2)$ ,复杂度较高。对这一问题的解决办法是:削减文法组合的量级,不对所有文法组合进行检索,而是以句法分析和语义角色标注为辅助,抽取论元对中的角色特征(包括施事短语、受事短语和谓词短语),以此类特征形成的两两组合形成查询。由此,文法组合的量级可由  $n^2$  降低为  $n$ ,相应查询次数也将大量削减,从而极大地提高系统效率。但是,这一方法是否会影响关系检测精度,还有待未来工作进一步分析研究。

### 9.3 健壮性分析

本文提出的平行推理方法对于信息检索有着较强的依赖性,详细来说,平行推理利用搜索引擎(Google)以及利用二元文法 PreGram 和 PosGram 构造的查询,通过信息检索过程,从 Web 数据中挖掘候选平行论元对,并在此基础上择优生成线索集合  $U_{key}$ ,然后进行后续推理计算。因此,检索性能的优劣将直接决定候选平行论元对的挖掘效果,进而影响  $U_{key}$  的质量和并行推理的精度。

然而,Web 数据是一种具有时新性的动态数据,换言之,Web 数据随时涌现大量新数据(如新闻),也因此,Web 数据内的特征、分布和关联等属性往往随时发生变化。基于这类属性的信息检索过程也将返回不同的结果,比

如,如果两条信息与查询具有近似的相关性,现有搜索引擎更倾向于将其中较新的信息作为检索结果予以输出(或将新信息的排序提前)。此外,搜索引擎是面向用户的信息服务平台,检索结果往往融入用户意图的分析和应用(如个性化信息检索),而不是纯粹基于文字语义学进行相关性计算(目前,语义计算刚刚在检索领域兴起),同时,搜索引擎附加了商业需求模式,针对特定需求,某些信息(如广告)强制出现且高位排序。总之,上述 Web 数据及搜索引擎的这类特性,往往造成检索结果质量的不确定性,而这一不确定性也必然影响检索驱动下的平行推理系统的健壮性。

为了检验平行关系推理系统的健壮性,实验对 JR+BPI, JR+PC+BPI 和 JR+FD+BPI 这 3 个主要推理系统进行了不同时间段的测试。详细说明,系统的测试时间不同,从 2012 年 8 月至该年 12 月,每个月对上述系统测试一次(表 8 与表 9 的测试结果发生于 2013 年 1 月),每次测试都使用当月的检索结果进行候选论元对挖掘,参数设置仍采用表 7 的训练指标(即  $n=2, N=20, window=10$ )。从而,各次测试的主要区别在于候选论元对集合及线索集  $U_{key}$  不同。测试统一采用 TD1 作测试集,不同时间段的测试结果见表 16。

**Table 16** Performances of parallelism inference systems in different months (test corpus: TD1)

表 16 不同月份的平行推理系统性能(测试集:TD1)

Systems	Aug.	Sep.	Oct.	Nov.	Dec.
JR+BPI (four-way)	0.433 0	0.392 9	0.559 2	0.453 2	0.533 4
JR+PC+BPI (four-way)	+3.34%	+5.26%	+1.82%	+3.63%	+2.86%
JR+FD+BPI (four-way)	+1.82%	+2.77%	+1.10%	+2.10%	+1.53%

结果显示,JP+BPI 系统的性能不稳定,最优时达到 55.92%的精度,最差时仅为 39.29%,相差 16.63%,说明检索质量的不确定性确实影响了推理的健壮性。不同的是,系统 JR+PC+BPI 和 JR+FD+BPI 都能给予 JP+BPI 系统相对稳定的性能提高,从而说明,基于线索集  $U_{key}$  的平行关系概率模型和基于同指焦点的消歧模型并未受到检索质量不确定性的直接影响,相对健壮。换言之,检索质量优劣只会影响基本的平行推理。

针对这一问题,一种解决途径是利用静态数据集作为检索和挖掘平行论元对的语言学资源,但为了确保能够获取足够多的平行论元对,需要这一静态语言学资源具有较大规模,并与测试语料(或实用时的目标语料)同属一种信息领域。换言之,语言学资源的静态性用于保证推理系统的稳定性,其规模大小将决定是否能够挖掘到平行论元对,而领域性则能保证挖掘所得的平行论元对质量(同领域的信息往往具有更高的语用习惯,平行性更易度量)。为了验证这一途径的可行性,实验额外使用了大规模静态新闻类数据 GIGAWORD,对系统 JR+FD+BPI 进行评测(系统使用 Apache Lucene 4.1 建立小型搜索引擎进行本地检索),获得如表 17 所示的性能,其中,加“\*”号标识的性能为 JR+FD+BPI 使用训练所得的参数设置进行测试的结果(即  $n=2, N=20, window=10$ ),未加“\*”号标识的性能为 JR+FD+BPI 在 GIGAWORD 上取得的最优性能,参数设置为  $n=2, N=16, window=10$ 。这一结果显示:使用大规模且同领域的本地知识库进行平行论元对的挖掘(注:PDTB 与 GIGAWORD 数据皆来自新闻领域),能够较好地支持基于平行推理机制的关系推理,性能接近利用通用搜索引擎和 Web 数据时的最优性能。同时,这一测试结果显示:平行推理机制在静态数据上的稳定性较好,对参数设置并不过分敏感,仅有  $N$  值发生小幅降低。 $N$  值的降低源于 GIGAWORD 的领域性较强,其语用形式更为收敛,有益于面向同为新闻领域的 PDTB 测试样本进行平行论元对的挖掘。换言之,给定特征论元对的二元文法组合( $n=2$ ),GIGAWORD 中蕴含相似文法的比例更高,参杂的噪声较少,从而选取较少的 TopN 值,即能较为准确地反映测试论元对的关系。

**Table 17** Inference performance using GIGAWORD for searching

表 17 利用 GIGAWORD 检索的推理性能

System	TD1	TD2
*JR+FD+BPI (four-way)	0.565 9	0.501 9
JR+FD+BPI (four-way)	0.569 7	0.512 6

## 10 总 结

本文针对隐式篇章关系检测任务,提出一种基于“语义平行则关系平行”的关系推理方法。这一推理方法通

过检索技术实现语义平行论元对的挖掘,并借助“显式指导隐式”的方式实现隐式关系推理.在此基础上,本文提出了关系平行概率模型和基于焦点同指的关系消歧,进一步优化这一平行关系推理方法,由此形成了推理系统,利用标准篇章关系语料库 PDTB 进行测试,其精度优于现有方法达 17.26%.

实验验证了句法结构特征在隐式关系分类中的劣势,并佐证了语义学特征对于识别和推理隐式关系的优势;同时,实验也验证了利用统计信息进行隐式关系推理的过程中,语言学资源的规模和类别的重要性;最后,实验验证了篇章关系分布规律的应用有助于提高关系平行概率的估计,以及论元焦点的识别与应用有助于歧义关系的发现与屏蔽.

然而,尽管平行关系推理方法在隐式关系检测的精度上优于现有方法,但关系检测的效率与健壮性却存在不足.平行推理效率不高的原因是,在检索驱动的平行论元挖掘过程中,查询的冗余度较高.针对这一问题,未来的工作尝试利用语义角色对应的文法构建查询,并尝试在提升效率的同时,利用语义角色对语义关系的约束力,提升平行推理的精度.健壮性不足的主要原因是 Web 数据的时新性和检索反馈的动态性,从而,借助数据挖掘技术和信息分类技术,自动建立大规模且领域一致的静态语言学资源并提升本地检索能力,是解决健壮性不足的必要途径,也是本文未来工作的重要尝试之一.

## References:

- [1] Miltsakaki E, Robaldo L, Lee A, Joshi A. Sense annotation in the Penn discourse treebank. *Lecture Notes in Computer Science*, 2008,4919:275–286. [doi: 10.1007/978-3-540-78135-6\_23]
- [2] Wolf F, Gibson E. Representing discourse coherence: A corpus-based analysis. In: *Proc. of the 20th Int'l Conf. on Computational Linguistics (COLING 2004)*. 2005. 134–140. [doi: 10.3115/1220355.1220375]
- [3] Marcu D, Echiabi A. An unsupervised approach to recognizing discourse relations. In: *Proc. of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)*. 2002. 368–375. [doi: 10.3115/1073083.1073145]
- [4] Wang WT, Su J, Tan CL. Kernel based discourse relation recognition with temporal ordering information. In: *Proc. of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*. 2010. 710–719.
- [5] Zhou MZ, Xu Y, Niu YZ, Lan M, Su J, Tan CL. Predicting discourse connectives for implicit discourse relation recognition. In: *Proc. of the 23th Int'l Conf. on Computational Linguistics (COLING 2010)*. 2010. 1507–1514.
- [6] Pitler E, Raghupathy M, Nenkova HM, Lee A, Joshi A. Easily identifiable discourse relations. In: *Proc. of the 22nd Int'l Conf. on Computational Linguistics (COLING 2008)*. 2008. 87–90.
- [7] Lin ZH, Kan MY, Ng HT. Recognizing implicit discourse relations in the Penn discourse treebank. In: *Proc. of the 2009 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2009)*. 2009. 343–351.
- [8] Soricut R, Marcu D. Sentence level discourse parsing using syntactic and lexical information. In: *Proc. of the 2003 Conf. of the North America Chapter of the Association for Computational Linguistics on Human Language Technology (NAACL 2003)*. 2003. 149–156. [doi: 10.3115/1073445.1073475]
- [9] Saito M, Yamamoto K, Sekine S. Using phrasal patterns to identify discourse relations. In: *Proc. of the Human Language Technology Conf. of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2006)*. 2006. 133–136.
- [10] Miltsakaki E, Dinesh N, Prasad R, Joshi A, Webber B. Experiments on sense annotations and sense disambiguation of discourse connectives. In: *Proc. of the 4th Workshop on Treebanks and Linguistic Theories*. 2005. 1–12.
- [11] Pitler E, Louis A, Nenkova A. Automatic sense prediction for implicit discourse relations in text. In: *Proc. of the Joint Conf. of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th Int'l Joint Conf. on Natural Language Processing of the Asian Federation (ACL-IJCNLP 2009)*. 2009. 683–691.
- [12] Nie JY, Simard M, Isabella P, Durand R. Cross-Language information retrieval based on parallel texts and automatic mining of parallel texts from the Web. In: *Proc. of the 22th Annual Int'l Association for Computing Machinery (ACM) Special Interest Group on Information Retrieval (SIGIR'99)*. 1999. 74–81. [doi: 10.1145/312624.312656]
- [13] Wu DK. Grammarless extraction of phrasal translation examples from parallel texts. In: *Proc. of the 6th Int'l Conf. on Theoretical and Methodological Issues in Machine Translation (TMI'95)*. 1995. 1–19.

- [14] Bao JP, Shen JY, Liu XD, Song QB. A survey on natural language text copy detection. Ruan Jian Xue Bao/Journal of software, 2003,14(10):1753-1760 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/1753.htm>
- [15] Meyer ADS, Garcia AAF, de Souza APD, Jr. de Souza CL. Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (*Zea mays* L). Journal of Genetics and Molecular Biology, 2004,27(1):83-91. [doi: 10.1590/S1415-47572004000100014]
- [16] Napoles C, Gormley M, Durme BV. Annotated gigaword. In: Proc. of the Joint Workshop on Automatic Knowledge Base Construction & Web-Scale Knowledge Extraction (AKBC-WEKEX) of NAACL-HLT. 2012. 95-100.
- [17] Wolf F, Gibson E. Representing discourse coherence: A corpus-based analysis. In: Proc. of the 20th Int'l Conf. on Computational Linguistics (COLING 2004). 2005. 134-140.
- [18] Prasad R, Dinesh N, Lee A, Miltsakaki E, Robaldo L, Joshi A, Webber B. The Penn discourse TreeBank 2.0. In: Proc. of the 6th Int'l Conf. on Language Resources and Evaluation (LREC 2008). 2008. 2961-2968.

## 附中文参考文献:

- [14] 鲍军鹏,沈钧毅,刘晓东,宋擒豹.自然语言文档复制检测研究综述.软件学报,2003,14(10):1753-1760. <http://www.jos.org.cn/1000-9825/14/1753.htm>



洪宇(1978-),男,黑龙江哈尔滨人,博士,副教授,CCF 会员,主要研究领域为篇章关系分析.  
E-mail: tianxianer@gmail.com



姚建民(1972-),男,博士,教授,CCF 会员,主要研究领域为机器翻译.  
E-mail: jmyao@szkj.gov.cn



严为绒(1990-),女,硕士生,CCF 学生会会员,主要研究领域为篇章关系.  
E-mail: sallyrong8521@gmail.com



朱巧明(1964-),男,教授,博士生导师,CCF 杰出会员,主要研究领域为中文信息处理.  
E-mail: qmzhu@suda.edu.cn



车婷婷(1989-),女,硕士生,CCF 学生会会员,主要研究领域为篇章关系.  
E-mail: chetingting1101@gmail.com



周国栋(1967-),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为信息抽取.  
E-mail: gdzhou@suda.edu.cn



梁颖红(1970-),女,博士,教授,CCF 会员,主要研究领域为智能信息处理.  
E-mail: liangyh7036@126.com