

电子科技大学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

学士学位论文

BACHELOR THESIS



论文题目 基于语法树的文档级关系预测算法设计
与验证

学科专业	计算机科学与技术
学 号	2020080902004
作者姓名	朱旭东
指导老师	康昭 副教授
学 院	计算机科学与工程学院（网络空间安全学院）

摘 要

文档级关系抽取旨在识别单个文档中实体对之间的关系。它需要处理多个句子并对这些句子进行推理。最先进的文档级关系抽取使用图形结构来连接文档中的实体，来捕获文档中的实体对的交互。但是这些方法没有充分利用在句子级关系抽取中被充分研究的语法信息。在本文中我们将语法树融合到文档级关系抽取中为主要研究内容，重点研究了使用依赖语法树，依存语法树进行文档级关系抽取算法的实现，以及怎么调整依赖语法树和依存语法树的在文档级关系抽取中的权重问题。我们利用依存语法树来聚合整个句子信息，并为目标实体对选择有指导意义的句子。同时我们利用依赖语法树对整个文档进行细粒度的分析，并选择其中重要的单词增强目标实体对的信息。文档级关系抽取将同时利用依赖语法树和依存语法树进行预测。通过在不同领域的数据集上的实验结果证明了该方法的有效性。

关键词：文档级关系抽取，依赖语法树，依存语法树，信息抽取

ABSTRACT

Document-level Relation Extraction (DocRE) aims to identify relation labels between entity pairs within a single document. It requires handling several sentences and reasoning over them. State-of-the-art DocRE methods use a graph structure to connect entities across the document to capture interaction between entity pairs in the document. However, this is insufficient to fully exploit the rich syntax information in the document, which is widely used in sentence-level Relation Extraction(RE). In this thesis, we focus on integrating syntax trees into DocRE as the main research topic, and investigate the effective and efficient implementation of DocRE algorithms using dependency syntax tree and constituency syntax tree, as well as how to adjust the weight of dependency syntax tree and constituency syntax tree in the extraction. It uses constituency syntax to aggregate the whole sentence information and select the instructive sentences for the pairs of targets. Meanwhile, it exploits the dependency syntax in a graph structure with constituency syntax enhancement and selects the most important words between entity pairs based on the dependency graph to enhance the information of target entity pairs. Finally, DocRE will integrate the dependency syntax and constituency syntax to predict. The experimental results on datasets from various domains demonstrate the effectiveness of the proposed method.

Keywords: Document-level Relation Extraction, Constituency Syntax, Dependency Syntax, Information Extraction

目 录

第一章 绪 论	1
1.1 研究的背景	1
1.2 研究的主要贡献与创新	2
1.3 研究的结构安排	3
第二章 关系抽取研究	4
2.1 关系抽取的定义	4
2.2 文档级关系抽取	4
2.2.1 基于序列的文档级关系抽取	4
2.2.2 基于注意力机制的文档级关系抽取	6
2.2.3 基于图的文档级关系抽取	9
2.3 本章小结	15
第三章 基于语法树的文档级关系抽取研究	16
3.1 单词嵌入	16
3.2 基于依存语法树的文档级关系抽取	16
3.3 基于依赖语法树的文档级关系抽取	19
3.4 融合两种不同语法树结果的文档级关系抽取	21
3.5 本章小结	22
第四章 基于语法树的文档级关系抽取实验结果	23
4.1 文档级关系抽取抽取数据集说明	23
4.2 DocRED 数据集实验结果说明与分析	24
4.3 DocRED 数据集实验结果说明与分析	24
4.4 医学数据集实验结果说明与分析	24
4.5 两种不同的语法树权重分析	26
4.6 消融实验结果说明与分析	27
4.7 依赖语法树路径距离分析	27
4.8 案例研究	28
4.9 本章小结	29
第五章 结 论	30
致 谢	31
参考文献	32

第一章 绪 论

1.1 研究的背景

关系提取是信息提取中的一项关键任务，旨在对非结构化文本中实体对之间的关系模式进行建模。它有两种特定的场景：句子级关系提取和文档级关系提取。与句子级关系提取 [1,2] 不同，文档级关系提取涉及识别和提取句子边界以外的实体之间的关系，为分析提供了更广泛的上下文，并且更具挑战性。这项任务可以从非结构化的大型文档（如科学论文、法律合同或新闻文章）[3] 中自动构建和填充知识库，以更好地理解实体之间的关系。因此，文档级关系提取更好地满足了实际需求，最近受到了越来越多的关注。

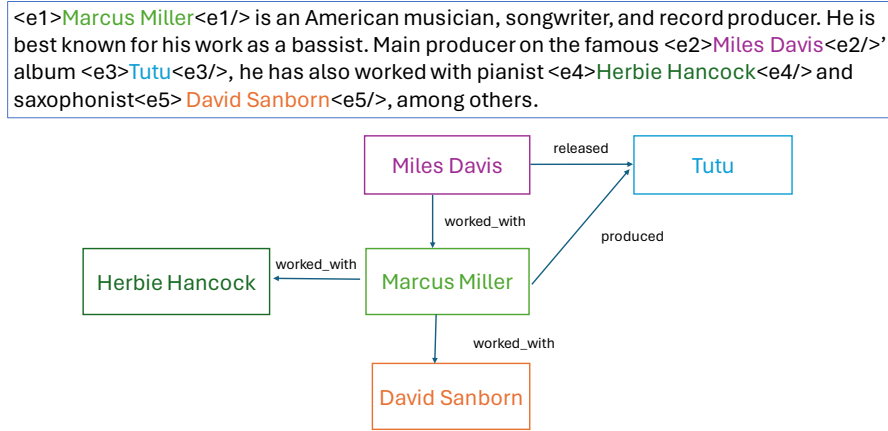
文档级关系抽取的目标是识别在同一文档内的实体对之间的关系。这是一项复杂的任务，因为它需要理解句子的内容并处理多个句子之间的交互。由于文档级关系抽取涉及多个句子，现在的研究有三个主要障碍 [4]：

1. 数量更多的潜在关系：相比于单个句子，一个文档中包含的实体数量更多。由于我们需要预测每两个实体之间的关系，因此潜在的待预测关系随着实体的增加呈指数级别增加。
2. 处理实体共指：一个实体通常在一个句子中只出现一次，而在一个文档中，它可以以多种形式出现多次。例如，在图1.1中的句子 2 和句子 3 中，“He”是“Marcus Miller”的共同指称。
3. 处理长距离关系：在文档级关系抽取中，我们通常需要预测跨句子之间的关系，这可能会导致实体对之间距离过长。然而，长句中通常包含不相关甚至有噪声的信息。*Huang* 等人 [5] 认为，在大多数情况下，最多只需要三个句子作为识别关系的支持证据。此外，这项工作指出，句子之间的重要性差异很大。不包含任何有价值的信息或关系的句子实际上阻碍了文档的理解。这也是我们这篇论文要解决的主要工作。

图 1.1 是一个例子，它包括来自同一文档中的句子级关系和文档级关系。为了推断“Marcus Miller”和“Herbie Hancock”之间的关系，模型应该能够排除不相关实体的影响，并找出句子 3 中的“he”一词指的是“Marcus Miller”。然而，由于现有的文档级关系提取模型 [6] 经常会被大量不相关的信息所淹没，从而无法捕捉到类似的关键信息。

由于以上这些原因，本文提出了新的一种文档级关系提取方法，通过融合语法信息作为辅助，来改进文档级关系提取。

图 1-1 文档级关系抽取的一个例子。图片上方是文档内容，文档中实体用不同的颜色标注，下方是文档级关系抽取的结果



1.2 研究的主要贡献与创新

本文主要处理文档级关系提取的长距离关系。由于长句中通常包含无关甚至嘈杂的信息 [7]，对于文档级关系提取来说，隐含地学习指导性上下文是不够的 [6]。最近的一些文章通过预训练模型和注意力机制来学习文档中复杂的交互 [5, 8, 9]，他们通过创建更复杂、更强大的变体 transformer 并利用注意力机制来隐含的捕获实体之间复杂的互动。但是，更多的方法通过显示捕获实体互动信息来改善文档级关系抽取。他们首先构建由不同节点组成的文档图（例如实体的提及节点，实体节点，句子节点或者文档节点）[10, 11]，将指导性上下文转化为图。由于语法信息可以通过提供显式语法细化和子内容建模来帮助文档级关系抽取 [12]，最近的研究 [13, 14] 采用依赖图来合并语法信息和结构上下文。他们发现一个精心设计的图形可以帮助模型更好的捕获实体互动信息，缩短实体之间的距离。然而，正如 [6, 15] 所指出的，尽管预训练模型是用大量真实世界的文本数据进行训练的，但隐式学习的语法与黄金语法之间仍然存在很大差距。事实上，语法信息在句子级关系抽取中得到了广泛的应用 [16, 17]，但在文档级关系抽取场景下尚未得到充分的探索。

为了充分利用文档中的语法信息，本文融合了依存语法和依赖语法。我们主要采用依赖图和依存树来合并额外的语法信息，并使用依存树中的信息来进一步增强依赖图的表示。依赖语法描述了一个句子中单词之间的依赖关系，这些依赖关系对原始纯文本有很强的补充作用。而依存语法可以分层合理地组织单个句子的不同单词，消除了枚举不同单词组合的过程，同时保持了分层语法信息。在树形深度学习模块 [12] 的帮助下，我们可以收集具有适当权重的所有子内容，以解

决句间依赖关系并融合依赖语法产生最终关系预测。

通过对三个公共文档级关系抽取数据集 (DocRED [18]、CDR [19] 和 GDA [20]) 的广泛实验下, 我们证明我们的算法在很大程度上优于现有方法。我们在这篇文章中的主要贡献可以总结如下:

1. 我们建议利用依存语法树来聚合句子级信息, 以弥补依赖语法树中的差距, 并通过添加文档节点来改进依赖图, 以减少实体对的距离, 简化长句交互。
2. 我们分别用树形长短期记忆模型和图卷积神经网络处理转化后的依赖语法和依存语法, 并设置一个可学习的参数来调整它们的权重。
3. 实验结果表明, 我们的模型在三个公共文档级关系抽取数据集上的测试优于现有方法, 特别是在 DocRED 数据集上, 我们的模型将最先进方法的 IgnF1 分数提高了至少 1.56%

1.3 研究的结构安排

本研究的后续方式组织: 第二章主要介绍文档级关系抽取的研究现状, 在这一章中, 我们对过往的研究结果进行总结。第三章介绍我们的研究工作, 包括我们依赖语法和依存语法的使用方式, 以及如何调节这两种语法的权重。第四章报告了我们研究的实验结果, 包括在三个公共文档级关系抽取数据集上的不同指标的分数, 与过往研究的分数对比以及分析; 同时包括了对我们方法的每个模块的分析。

第二章 关系抽取研究

在这一部分，我们首先对关系抽取进行定义。然后我们介绍对文档级关系方法的回顾。

2.1 关系抽取的定义

在这篇文章中，我们遵循 Zhang 等人提出的定义 [21]。对于每个预测任务，我们有文档 $D = \{S_i\}_{i=1}^{n_s}$ 和实体集合 $E = \{e_i\}_{i=1}^{n_e}$ 。其中 n_s 是文档中句子的数量， n_e 是实体的数量，在句子级关系抽取中 n_s 被固定为 1，但是在文档级关系抽取中没有限制。文档中的句子被表示为 $S_i = \{w_j\}_{j=1}^{n_w^i}$ ，其中 w_i 是第 i 个句子中的第 j 个单词。每一个实体被表示为 $E_i = \{m_j\}_{j=1}^{n_m^i}$ ，其中 m_i 是第 i 个实体中的第 j 个提及。在句子级别设置中，一个实体只由一个提及表示，即 $\forall i \in [1, n_e], e_i = m_i$ ，但是在文档级关系抽取中情况并非如此，其中实体可以通过多次提及来表示。为了获得实体的固定表示，以前的方法首先对所有提及的嵌入进行平均，即 $e_i = \frac{\sum_{j=1}^{n_m^i} m_j}{n_m^i}$ ，或者对所有提及的嵌入进行最大池化。然而，现在的方法都使用更加柔和的 logsumexp 表示，即 $e_i = \log \sum_{j=1}^{n_m^i} \exp(m_j)$ ，这样可以促进来自弱提及信息的积累 [22]。我们也采用这种方式表示每一个实体。

文档级关系抽取的目标是正确推断每个实体对 $(e_s, e_o)_{s,o=1,2,\dots,n_e; s \neq o}$ 之间的关系。其中 e_s 是头实体， e_o 是尾实体。预测的关系是预定义的关系集合 R 和 NA （没有关系）的子集。

2.2 文档级关系抽取

本节介绍了文档级关系抽取的最新进展，并将现有的文档级关系抽取方法大体分为三类：基于序列的，基于注意力机制的和基于图的文档级关系抽取方法，并分析他们的优缺点。

2.2.1 基于序列的文档级关系抽取

执行文档级关系提取的一种方法是将文档视为一个长的增强的序列，并应用从句子级关系抽取导出的序列模型来识别特定实体之间的关系。

2.2.1.1 基于卷积神经网络的文档级关系抽取

Gu 等人 [23] 通过两个分离的模型来提取关系。他们首先将最大熵系统用于句子间的预测，卷积神经网络模型用于句子内的预测。然后，这两个结果被合并以获得文档级别的关系。此外，他们还定义了一些启发式规则来从摘要中提取关系，例如将论文标题中提到的所有化学物质（或如有必要，将摘要中最频繁的化学物质）与摘要中提及的所有疾病相关联。

李等人 [24] 使用递归分段卷积神经网络进行化关系提取，并结合了领域知识、分段策略、注意力机制和多实例学习。该方法首先利用分段卷积神经网络来捕获单个实例的表示。然后，使用递归神经网络来聚合实例的表示，从而能够为实体对形成全面的文档级表示。

2.2.1.2 结合卷积神经网络和递归神经网络的文档级关系抽取

Mandya 等人 [25] 提出了一种长短期记忆和卷积神经网络的组合模型，该模型利用单词嵌入和位置嵌入来提取跨句子的 n 元关系。长短期记忆模型将结合单词嵌入和位置特征的变换矢量表示作为输入，将输出被馈送到卷积神经网络中。卷积神经网络被用来提取最重要的特征。这种方法的优点是独立于复杂的句法特征，如依赖树、共指消解或话语特征。因此被广泛应用到后续的改进模型中。

为了提取基因-疾病相关性，Wu 等人 [20] 还在其方法中结合了门控神经单元。在这里，卷积神经网络首先使用具有不同宽度的不同滤波器从单词表示中计算句子表示，以捕捉不同的特征。然后，门控神经单元将它们转换为文档表示。通过消融实验，他们发现，即使单独的卷积神经网络也可能捕捉到文档的一些重要特征，而额外的门控神经单元层可以提高其性能。这项研究后来被 Su 等人 [26] 进行了扩展。在 Wu 等人的系统之前引入了部分过滤步骤，来删除不包含任何实体对信息和模糊关系的段落。

2.2.1.3 以实体为中心的文档级关系抽取

在传统的句子级关系抽取中，对提及元组进行分类预测就已经足够了，但是在文档级关系抽取中却不是这样。因为任何实体都可能有多次提及，应用简单的提及级别分类方法会使模型在大量的提及中淹没，导致预测失败。为了解决这个问题，Jia 等人 [22] 提出通过对每个实体元组进行单一关系预测，使用以实体为中心的多尺度预测。它们通过对实体的所有上述表示应用 logsumexp 来获得实体表示。

2.2.1.4 基于序列的文档级关系抽取总结

这些方法通过将文档视为一个长的增强的序列，将句子级关系抽取的方法自然的映射到文档级关系抽取上。这些方法的优点是能够处理长文档，并且能够利用跨句子的关系信息。然而，当待预测实体对涉及到处理长距离并需要多跳推理上下文内容时，这些模型不可避免地会遇到挑战。

2.2.2 基于注意力机制的文档级关系抽取

2.2.2.1 早期基于注意力机制的文档级关系抽取

Verga 等人 [27] 引入了双仿射关系注意力网络来计算实体之间的关系，它同时预测实体之间的句子内关系和句子间关系。为了实现这一点，它首先使用 transformer 层对令牌嵌入进行编码，然后它使用双仿射运算产生实体提及对之间的预测。最后，它使用 logsumexp 来汇集每个实体提及级别预测，以获得实体级别预测。该系统可以针对命名实体识别和关系预测进行联合训练，这可以提高对噪声的鲁棒性；此外，该系统在编码长序列方面相比于以往的方法更有效。

为了缓解正负关系之间的数据不平衡问题，Wang 等人 [28] 采用了两步微调过程和预训练模型。前者识别两次提及之间是否存在关系，后者仅使用先前识别的关系事实来训练模型来识别这些实体。

为了增强语言表示模型中的共指推理，Ye 等人 [8] 在各种下游任务上测试了他们的系统，这其中也包括文档级关系抽取。他们引入了一种称为提及参考预测的新的预训练任务。该任务利用了文章中的重复提及，并采用提及参考掩蔽来预测相应的参考，而不是依赖于监督的共同参考数据进行微调。此外，他们引入了基于新的训练目标，以鼓励基于上下文的单词选择和上下文敏感性的表示，从而促进共指推理。

2.2.2.2 分层注意力模型

为了将推理信息从实体级聚合到句子级，再聚合到文档级，Tang 等人 [29] 使用层次推理网络进行预测。受 Bordes 等人 [30] 中给出的平移约束的启发，该约束对关系三元组 (r, e_s, e_o) 进行了建模。他们通过将实体提及和关系联合建模，获得了实体级别的表示信息。然后，他们应用语义匹配方法将实体级推理信息与每个句子向量进行比较，并使用多层双向长短期记忆模型获得句子级信息。最后，他们重新利用注意力机制来识别关键的句子级推理信息，这有助于以更全面的方式全面表征文档级推理。虽然这种方法十分有效，但是它的正确性受到广泛质疑，因为在文档级抽取中，两个实体对可以具有多个关系，一个实体可以与多个其他实

体具有相同的关系。

Kuang 等人 [31] 认为，实体对之间的关系通常可以通过几个关键词来推断。为了从中受益，他们使用自注意记忆模块与预训练模型结合，通过寻找具有高交叉注意力实体对的单词嵌入来捕捉关键词特征。他们的模型非常简单，在领域和通用数据集上都取得了很好的改进，并可应用于其他数据集。

2.2.2.3 基于实体表示的注意力模型

Han 和 Wang [4] 利用具有类型信息的文档级实体掩码方法来提供更多关于文档级关系预测中实体的信息。它在预训练模型的嵌入中引入了实体特征，用上下文信息和实体特定信息丰富了实体的表示。Han 和 Wang 首先使用预训练模型来获得实体提及的上下文化表示。然后，它使用一次性关系预测方法一次性处理所有实体对，因此文本只编码一次。

采用与 Jia 等人 [22] 相同的实体中心主义观点，Yu 等人 [32] 还解释了每一次提及之间的重要性，以代表关于特定关系的实体。他们使用关系特定提及注意力网络 (RSMAN) 将每个关系的基本语义合并到关系表示中，然后计算候选关系和给定实体对的提及表示之间的注意力。通过考虑这些事项，该模型通过对所有提及的表示进行加权聚合来获得实体的合成表示。这使得模型能够在不同的表示空间中捕获来自多个提及的信息，从而为不同的候选关系产生灵活的、特定于关系的实体表示。作为一种插入式方法，RSMAN 可以用作任何关系预测模型的方法中。

如前所述，在文档级关系预测中，一个文档包含多个实体对，这些实体对可以在文档中多次出现，并具有多种可能的关系。为了解决这些问题，Zhou 等人 [33] 在预训练模型的基础上使用了两种技术：自适应阈值和本地化文本池 (ATLOP)。自适应阈值通过将多标签分类的全局阈值与可学习的依赖于实体的阈值进行切换来为每个实体对确定最佳阈值。而本地化文本池直接将注意力从预先训练的语言模型转移到有助于确定关系的相关上下文。这解决了对所有实体对使用相同实体嵌入的问题。ATLOP 通过与当前实体对相关的额外上下文丰富了实体嵌入，实现了更全面的表示，并为 DocRE 带来了巨大的飞跃。在我们的方法中，我们也采用了它的这两种技术。

2.2.2.4 基于远程监督标签的注意力模型

Xiao 等人 [34] 试图通过三个预训练任务来减少远程监督标签中引起的固有噪声：提及实体匹配，关系检测和关系事实对齐。提及实体匹配涉及文档内和文档间的匹配，第一种侧重于在文档内建立共引用连接，而另一种则捕获文档之间的实体关联；关系检测，主要检测被错误标记为负面的正面关系；关系事实对齐确

保模型即使以不同的方式表示，仍然为同一实体对生成一致的表示，

接着，Tan 等人 [35] 提出了一种利用远程监督数据的多步骤方法。首先，他们使用注意力模块作为特征提取器，允许关注逻辑路径中的不同元素，并捕捉关系三元组（关系，头实体和尾实体）之间的相互依赖性。其次，为了处理类的不平衡，他们引入了一种关注长尾类的自适应焦点损失，使它们对整体损失的贡献更大。最后，他们利用知识提炼来弥合手动注释数据和远程监督数据之间的差距。它由一个用少量人工注释数据训练的教师模型组成，该模型对大量远程监督数据生成预测，从而为预训练学生模型提供软标签。

2.2.2.5 基于结构推理的注意力模型

Xu 等人 [36] 基于当时文档级关系抽取模型的一个问题，即上下文推理和结构推理的阶段是分开的、因此缺乏对上下文表示的结构指导，从而引入了结构化自注意网络（SSAN），将结构依赖性包含在编码网络内和整个系统中。他们利用自注意机制的一个新扩展 [37]，通过考虑两个实体结构，一个用于同一句子中的实体对的共现，另一个用于单个实体的共指，有效地对其构建块内和所有网络层自下而上的提及依赖性进行建模。除了实体提及之间的依赖性，Xu 等人 [36] 还使用注意力模块来捕捉实体提及与句内非实体之间的关系。它们包含了一个额外的与注意力模块并行的模块，该模块以上下文化的查询键表示为条件，对结构依赖性进行建模。这种设计使模型能够受益于结构相关性的指导，确保上下文信息和结构信息有效集成。

2.2.2.6 基于联合学习的注意力模型

Ma 等人 [38] 解决了学习文档级关系抽取和证据提取的两个问题：高资源消耗和缺乏人工注释数据。他们使用证据引导注意机制进行的文档级关系提取，通过在证据提取上指导 ATLOP 模型 [9] 来完成证据提取和节省资源消耗。Ma 等人使用手动标注的实体对特定的局部上下文嵌入进行计算，并将注意力权重作为证据得分结合在标记上。这种方法允许他们引入来自证据提取的知识，而不引入额外的可训练参数或昂贵的矩阵计算。为了缓解数据的缺乏，他们使用了与 Tan 等人 [35] 类似的知识提取技术，利用远程监督数据来训练学生模型。

2.2.2.7 基于注意力的文档级关系抽取模型小结

综上所述，基于注意力的文档级关系抽取模型的研究已经取得了长足的进步。目前，除了早期的一些朴素方法，现有的方法大致可以分为以上五类，这些方法可以将实体嵌入、关系嵌入和上下文信息结合起来，以获得比基于序列的模型更

好的文档级关系抽取结果。这些模型的共同特点是使用注意力机制来捕捉不同实体和关系之间的依赖关系，并将它们融合到统一的表示中。它们的优点在于能够捕捉到实体和关系之间的复杂关系，并能够在不同的表示空间中捕获来自多个提及的信息。这些模型的缺点是需要它们要么需要额外的远程监督数据，要么需要设计复杂的分层结构。而且由于实体之间的距离可能较长，它们中大多数方法没有办法有效的处理长距离交互。然而，它们中的一些思想和假设十分有效（如第2.2.2.3节的ALTOP的自适应阈值和本地化文本池等），因此在最近的研究中被广泛采用。

2.2.3 基于图的文档级关系抽取

在这一节中，我们首先在第2.2.3.1节介绍早期的一些图的文档级关系抽取方法。然后我们将现有的基于图的文档级关系抽取模型分为五类：分别在第2.2.3.2节中介绍了联合学习的图文档级关系抽取方法，在第2.2.3.3节中介绍了以实体为中心的图文档级关系抽取方法，在第2.2.3.4节中介绍了基于异构图的图文档级关系抽取方法，在第2.2.3.5节中介绍了基于分层网络的图文档级关系抽取方法，在第2.2.3.6节中介绍了基于最优图结构的图文档级关系抽取方法。

2.2.3.1 早期的图文档级关系抽取方法

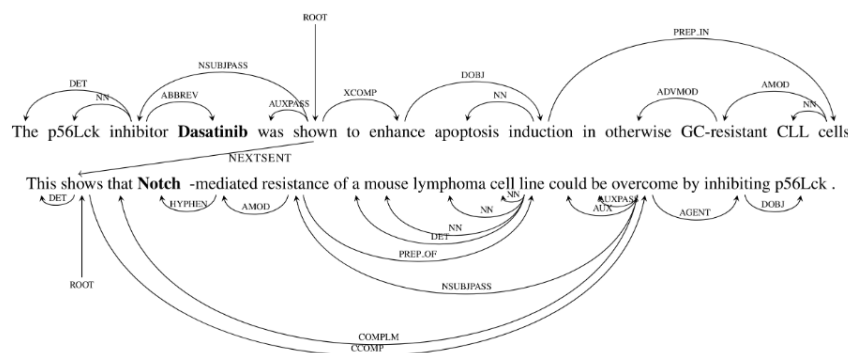


图 2-1 Quirk 和 Poon [39] 构建的文档图

Quirk 和 Poon [39] 引入了文档图的概念。它提供了一种统一的方法，通过将单词表示为节点，将一些启发式的关系（例如，依赖关系，邻接关系等）表示为边，来导出对待预测的关系进行分类的特征，图2.2.3.1是这种文档图的一个示例。这为文档级关系抽取的所有基于图的方法铺平了道路。Peng 等人 [40] 决定将这种图表示作为其框架的主干，其中包含一种新的图形长短期记忆模型。在经典的长短期记忆模型体系结构中，每个单元只有一个前置点（因此只有一个遗忘门）。图

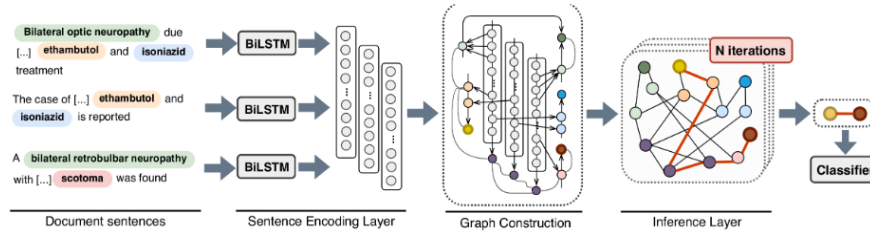


图 2-2 Christopoulou [42] 等人用于句子级关系抽取的基于图的随机行走的神经网络。该模型对句子进行编码，并将其输入迭代算法，以生成目标实体节点之间的边缘表示。为简洁起见，未显示某些节点连接

形长短期记忆模型的情况并非如此，通过从树形长短期记忆模型 [41] 中获得灵感，它可以具有多个前置点，包括通过不同边与同一单词连接的点，并为每个单元添加与先例数量一样多的遗忘门。从输入文本获得的单词嵌入被作为图形长短期记忆模型的输入，该图捕捉每个单词的上下文表示。然后，实体的上下文表示通过这种图形的连接进行组合，并由关系分类器使用。在多单词实体的情况下，Peng 等人计算其单词表示的平均值来获得实体的嵌入向量，尽管现在大多采用 logsumexp 操作来获得更好的结果，但这种方法在当时是首创。至于图形长短期记忆模型的反向传播，他们将文档图拆分为两个有向无环图。一个有向无环图包含从左到右的线性连接链接相邻的单词以及其他前向指向依赖项（前向传递），而另一个有向无环图则涵盖从右到左的线性连接链接相邻的单词和后向指向依赖性（后向传递）

Christopoulou 等人 [42] 提出了一种仅用于句子级关系抽取的基于图的随机行走的神经网络，同时处理多对关系，并将句子表示为向无环图，这种方法的详细过程参见图2.2.3.1。他们后来在将他们的方法扩展到文档级别，引入部分连接的文档图而不是以前的完全连接的文档，使得文档级关系抽取向前迈出了一大步。与仅包含实体节点和单个边缘类型的句子级图相比，这种新的文档图由异构类型的节点和边组成，当实体提及可用时，这种方法可以使用多实例学习来进一步提高模型的性能。节点之间的连接基于预定义的文档级交互，目标是使用图中的其他现有边生成实体到实体的边表示，从而推断实体到实体关系。在这项任务中，Christopoulou 等人 [43] 还发现，文档级信息有助于找到句内关系。同年年初，同一团队 [13] 中提出应用图卷积网络实现消息传递。在该方法中，利用依赖语法边、相邻句子和单词边、自节点边以及共引用边来构造文档图，以连接句子的依赖树。Guo 等人 [44] 还提出通过注意力引导来扩展关系抽取的图卷积网络结构，以有效地关注任务的相关子图，而不修剪一些重要信息。

根据 Quirk 和 Poon [39] 中引入的依赖图，Gupta 等人 [45] 捕捉了两个实体之

间的最短依赖路径，同时考虑了句内和句间关系。它们使用双向循环神经网络来表示通过依赖子树的最短依赖路径上的每个单词。他们提出的方法，基于句子间依赖的神经网络，通过将句子解析树的根节点连接到后续树的根，结合最短依赖路径来处理句子间关系。此外，他们采用了一种增强依赖路径，其中双向循环神经网络对句子间最短依赖路径上每个单词的依赖子树进行建模，然后提取显著的语义特征。在大多数情况下，增强依赖路径模型比最短依赖路径模型更好，然而，他们没有在 DocRE 的基准数据集上进行测试。当将他们的方法与 [40] 进行比较时，他们指出，由于其非循环结构架构，他们的系统更简单、更高效。

此外，Minh Tran 等人 [46] 发现基于图的过往模型只关注图的边缘表示，而忽略了节点的表示。因此，这些模型错过了实体及其提及中包含的潜在重要线索，并阻止模型利用节点和边缘表示之间的关系，以及利用实体的表示向量之间的相似性找到相同实体的节点。因此，他们扩展了 Christopoulou 等人 [43] 的 EoG 模型。他们通过合并三个损失来提高模型表现：一个用于节点表示，一个用于结点边缘表示一致性，以及一个用于实体提及表示一致性。

2.2.3.2 基于联合学习的图文档级关系抽取

Luan 等人的工作 [47] 主要集中在联合学习中执行不同的信息抽取任务。他们首先提出了一个句子级信息抽取系统，使用动态跨度图同时执行命名实体识别，关系抽取和共指消解。由于模型可以随着新信息添加到输入文本中而更新，因此它能够传播实体共指和关系类型置信度，这有助于跨度表示的迭代细化。这与以前的模型形成了对比。在以前的模型中，任务之间的唯一交互发生在共享的第一层长短期记忆模型中。

Wang 等人 [48] 使用一种新颖的文法结构理论来建立有效的语义联想，以及一套证据推理方法来选择合适的证据，以此来训练他们的话语感知神经关系抽取模型。通过将关系抽取和证据提取进行联合学习，这种方法缓解对关系抽取的证据缺乏选择和推理能力的问题。

在类似的方向上，Xie 等人 [49] 提出通过有效的证据提取和推理阶段融合来增强文档级关系抽取的能力。在训练步骤中，对模型进行关系抽取和证据提取的联合训练，而在推理过程中，对原始文档和提取证据的预测结果进行组合，这鼓励模型专注于重要句子，同时减少信息损失，强迫模型在句子中进行更多推理，使其能够包含更多关于关系的知识，从而使模型具有更高的性能。最近的一些工作 [12, 14]，将文档图和证据提取结合到一起，利用语法信息来辅助进行文档图的构建和证据的提取。但是他们将两者孤立的进行训练，而没有考虑到它们的相互融合。

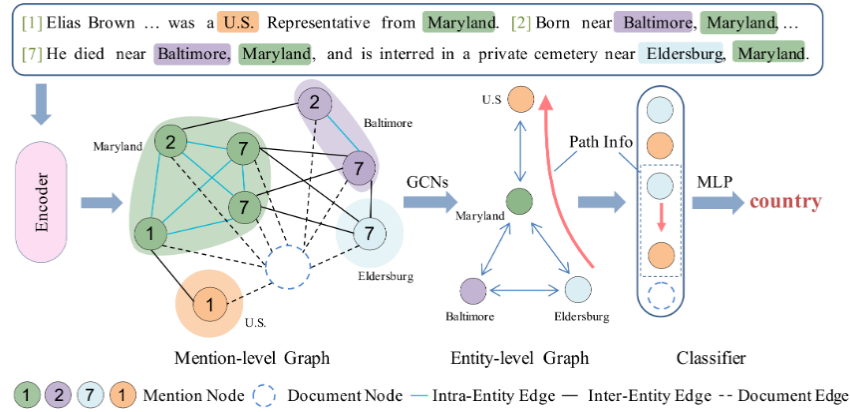


图 2-3 图聚合和推理网络的结构 [10]。首先，上下文编码器获得每个单词的上下文化表示。然后，利用提及节点和文档节点构建提及级图。在应用图卷积网络之后，构建了实体级图，其中识别实体之间的路径以进行推理。最后，分类模块利用收集的信息对目标关系进行预测。不同的实体有不同的颜色

2.2.3.3 以实体为中心的图文档级关系抽取

与以实体为中心的注意力模型相似，基于图的文档级关系抽取模型也对实体信息做了扩展。Zeng 等人 [10] 提出了图聚合和推理网络，该网络结合了由提及级别和实体级图组成的双图设计。图聚合和推理网络捕捉不同提及之间的复杂互动，并聚合属于同一实体的互动。这是通过三种类型的边来实现的：一种是执行共指解析的实体内边，一种是对句子中实体之间的交互进行建模的实体间边，以及一种文档边、所有提及节点都连接到文档节点。通过建立这些关系，文档节点可以有效地处理所有提及，并促进文档和单个提及之间的交互。它还解决了长距离依赖的挑战，因为文档节点可以作为使用 GCN 传递消息的枢纽（见图 2.2.3.3）。但是这种方法也有一个显著的缺点，那就是它的计算复杂度过高，无法处理大规模数据集。

为了考虑共指信息，Xue 等人 [8] 提出了基于图推理网络的文档级关系抽取（CorefDRE），通过提及代词聚类直接指示模型学习共指信息。他们使用提取所有提及代词对，并使用预训练模型计算每个提及代词对的可信度，以减少代词带来的噪声。CorefDRE 首先利用提及代词对及其可信度得分，结合代词节点、提及节点和各种类型的边，构建提及代词可信图。然后，它将图卷积网络应用于提及代词可信图，以捕获提及和代词的有意义的表示，同时也抑制了可能的噪音，将其转化为同构实体图。最后，模型聚合图中的节点和边表示，获得用于有效分析和

理解的综合表示。

与 Xue 等人 [8] 类似, Li 等人 [50] 介绍了一种基于自监督仿射计算对单词和实体进行聚类的共指图。这些聚类作为一个节点阵列, 能够捕获跨句推理的线索, 然后将它们转换为图。该图将共指图与 Zeng 等人 [10] 提出的提及图相结合, 使其成为实体图。这种模型采用了关系图卷积网络和噪声抑制机制 [8], 产生了同构实体图, 这使得即使这种方法允许增强长距离关系的建模, 两个图的合并仍然不会引入不希望的噪声。

2.2.3.4 基于异构图的图文档级关系抽取

基于基于异构图的图文档级关系抽取与基于分层网络的方法有些相似, 但是基于异构图的方法更强调使用不同的图形, 而不是使用分层网络来辅助文档级关系抽取。Zhang 等人 [21] 提出了第一种双层异构图, 有效地将这顺序信息和结构信息结合起来, 以促进多跳推理和最终预测。该模型主要由一个结构建模层组成, 考虑了对文档固有结构进行全面编码, 包括文档中单词的顺序、语法和层次结构。关系推理层通过在不同实体之间传播关系信息来促进多跳关系推理。

然而, 这种模型中的大多数构建了所有实体对之间的关系信息, 而不知道它们之间是否存在关系。因此, Xu 等人 [51] 提出使用重构来迫使他们的基于异构的图自注意网络更加关注具有标注关系的实体对的学习。重构器沿着连接两个实体节点的路径生成一系列节点表示, 当实体对之间存在标注关系时, 最大化路径的概率。这样可以进一步提高模型的性能。

借鉴前人的工作, Wang 等人 [21] 提出了一种多层异构图神经网络模型。它构造了一个多层异构图, 包括单词级图、提及级图和实体级图。前两个图分别使用面向节点类型和边缘类型的异构图注意力网络进行消息传递。然后, 实体级图采用基于门控的路径推理机制来提取多跳关系。这种方式和我们介绍的基于分层网络的方法就非常相似了。

2.2.3.5 基于分层网络的图文档级关系抽取

为了区分句内推理和句间推理, Kim 等人 [52] 在其全局关系提取器中使用了一种分而治之的方法, 通过使用两个独立的模型来提取局部和全局关系并构建知识图来完全避免这个问题。全局关系提取器由一个基于长短期记忆的句子级关系抽取模块和一个负责从局部关系生成知识图的知识图构造器模块组成; 两个模型都是交替训练的。

Li 等人 [53] 采用基于局部和全局提及的推理网络来区分近距离和远距离实体提及的影响。他们添加了一个新的共同预测器, 利用局部和全局实体表示来联合

推理近实体和远距离实体之间的关系。该模型包括一个二维窗口卷积，用于捕捉关系的实体对之间的局部提及的交互。此外，模型还添加了一个注意力模块，用于捕捉每对实体提及之间的全局交互，这些实体提及彼此交互以提供局部和全局上下文之间的信息。该模型获得了很高的性能，证明了句内关系和句间关系之间的相互作用对文档级关系抽取的模型的有效性。

在类似的思想下，Zeng 等人 [54] 强调了文档级关系抽取中处理句内关系和句间关系的常见做法。由于二者高度相似，这导致在预测这些不同关系类型时缺乏区分性。而且并非所有的实体对都可以通过路径连接，一些逻辑推理路径可能无法完全捕获在图中，这导致在许多实体对无法被逻辑推理覆盖。他们的方法将句内推理和句间推理分开。对于句内关系，他们使用句子级编码器来表示每一个同时出现的实体。通过聚合所有共现实体的关系表示，它们获得了一个最终表示，该表示鼓励句内实体对在其特定上下文中强调局部模式。为了捕捉句子间的关系，继他们之前的工作 [10] 之后，他们采用了文档级编码器和提及级图，使模型能够捕捉全局和局部上下文。为了选择性地识别可能的跨句关系的句子，他们使用证据选择器，它鼓励句子间实体对选择性地关注为其跨句关系提供支持证据的句子。

2.2.3.6 基于最优图结构的图文档级关系抽取

Nan 等人 [55] 发现使用共同引用和规则构建图并不总是导致最优的图结构。因此，他们建议将图结构视为一个潜在变量，并以端到端的方式对其进行归纳，以增强整个文档中的信息聚合。他们的潜在结构优化模型由一个节点构造函数和一个动态推理器组成，节点构造函数将节点作为每次提及的上下文表示，动态推理器从迭代优化的节点中生成文档级结构。

Huang 等人 [5] 专注于研究在流行数据集中识别关系需要多少句子。他们发现，在 DocRED 数据集中，当段落平均由八句话组成时，96.1% 的例子需要不超过三句话作为支持证据，87.7% 的例子甚至只需要两句或更少。CDR 和 GDA 数据集也进行了类似的计算。研究人员给出的一种可能的解释是，考虑到人类工作记忆的局限性，关系抽取任务的性质自然局限于有限数量的实体和上下文信息。正如 Barreyro 等人 [56] 所述，工作记忆在推理任务中尤为重要，但其分离信息块的能力通常限制在四句话左右。因此，在数据中观察到的三句话信息块的共同模式可能是我们用于相互理解的潜意识策略。利用这些假设，Huang 等人设计了三个启发式规则来从每个文档中提取路径，旨在近似支持证据：连续路径（头和尾实体在三个连续的句子中）、多跳路径（头实体和尾实体分别相隔在长距离的句子中，但可以通过与不同句子中的和头实体和尾实体同时出现的枢轴实体进行连接）和默认路径（所有成对的句子，其中一个包含头实体，另一个包含尾实体）。即使他

们的方法看起来过于简单，但当论文发表时，他们通过将其与双向长短记忆模型相结合，比一些最先进的方法获得了更好的结果 [10,55]，这表明明确选择最佳路径比完全依赖基于图的模型要好。

2.2.3.7 基于图的文档级关系抽取总结

基于图的文档级关系抽取方法有着广泛的应用，他们相比于基于注意力的方法有着明显的优势，因为它们可以将关系推理的结构信息结合到图结构中，从而更好地捕捉长距离关系。然而，这些方法要么使用启发式的图结构，无法捕获原文中的顺序信息 [10,51]；要么分别使用依赖性和依存信息捕捉不同粒度的信息来预测不同的任务 [12,14]，但没有将它们结合到一起。在我们这项研究中，我们通过结合依存和依赖信息，并使用依存树增强依赖图来克服以上缺点。通过引入不同的语法，我们的模型可以融合不同粒度的信息，更好地捕捉长距离关系。

2.3 本章小结

在这一章中，我们深入探讨了围绕文档级关系抽取任务的现有文献的大体方法和他们对应的优缺点。除了传统的基于序列的方法，我们将基于注意力的方法和基于图的方法各自分为五种。可以观察到，这两类方法有很多相通之处，利于他们都重视实体信息，各自提出一些假设来丰富实体信息。再者二者都重视文档的结构信息，都利用结构信息来辅助推理。但是，基于图的方法不仅有更优秀的性能，也有着更好的解释性。通过利用语法信息将文档转化为文档图，我们可以更好的捕捉实体对之间的关系，同时缩短实体对之间的距离，简化它们之间交互的同时也能保存足够多的信息。因此本研究立足于图形的方法来改进文档级关系抽取任务。

第三章 基于语法树的文档级关系抽取研究

为了充分利用文档中的语法信息，本文融合了依存语法和依赖语法来改进文档级关系抽取。我们主要采用依赖图和依存树来合并额外的语法信息，并使用依存树中的信息来进一步增强依赖图的表示。依赖语法和依存语法描述了语法信息的互补但不同的方面。图 3.3 中的依赖图主要用于整合单个句子内的句法信息，该信息与原始纯文本具有很强的互补性，而图 3.2 中的依存树则分层合理地组织单个句子的不同单词。

我们算法的总体架构如图 3.1 所示。我们利用依赖语法和依存语法来构建依赖图和依存树，并利用预训练模型对文档中的单词进行编码。然后，我们使用树形长短期记忆模型来聚合来自依存树的信息，并通过图神经网络利用依赖图的信息，同时利用依存树来改进依赖图。我们通过可学习权重来调节依赖图和依存树的关系，以更好的利用这两种语法信息。

3.1 单词嵌入

给定一份文档 D ，我们在每次实体提及之前和之后都会首先插入一个特殊的标记“*” [57]。然后，我们将来自文档 D 的单词馈送到预训练模型中，以获得上下文化表示 $\mathcal{H} = \{h_1, \dots, h_T\} \in \mathcal{R}^{T \times d}$ ，其中 d 是训练模型嵌入后令牌的数量， d 是令牌嵌入的维度：

$$\mathcal{H} = PLM\{x_{1,1}, \dots, x_{L,P_l}\} \quad (3-1)$$

$x_{i,j}$ 是文档中 i^{th} 句子的 j^{th} 单词，PLM 可以是预训练的 BERT 模型 [58] 或长短期记忆模型。

3.2 基于依存语法树的文档级关系抽取

在通过预训练模型获得令牌级嵌入之后，我们的模型利用依存语法树来获得分层语法信息，来补偿过往模型预训练的依赖语法树中的不精确性。具体来说，我们利用基于文档 \mathcal{D} 的依存树，并使用树形长短期记忆模型 [12, 59] 来合并每一个

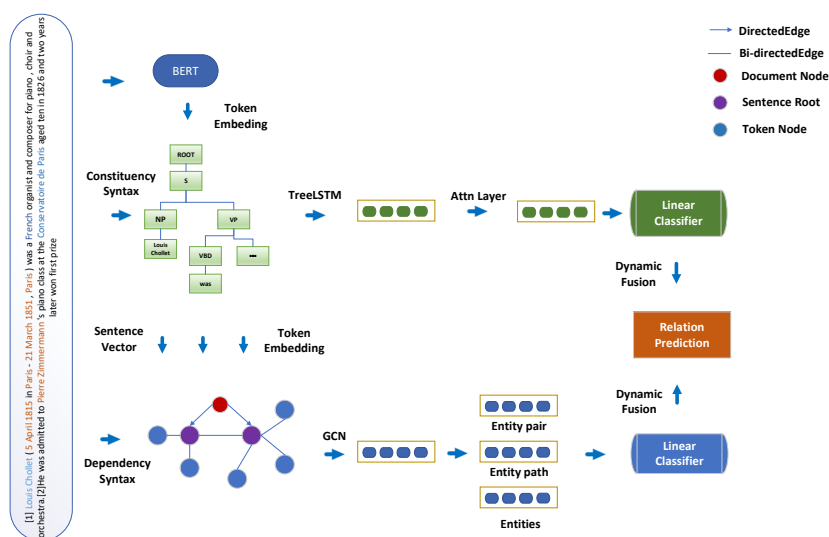


图 3-1 整体算法的概述。我们使用依存语法的结果来增强依赖图，并通过动态加权融合获得实体对之间的关系

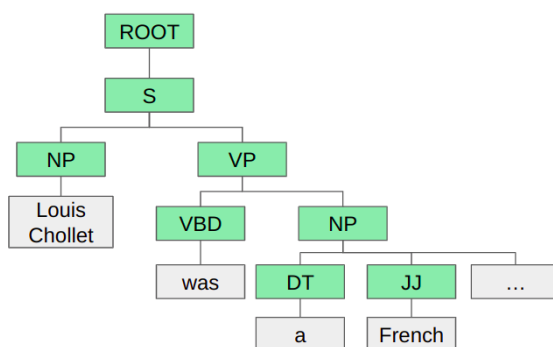


图 3-2 依存树将句子组织成树结构，不仅可以引入额外的层次语法信息，还可以在任意粒度上探索子内容。非叶节点颜色为绿色

句子的依存树根节点的信息。首先，我们将每个句子解析为相应的依存树^①。我们可以从图3.2中看到，每个依存树都描述了一种分层表示整个句子的语法结构方法。使用依存树，我们不仅可以包含额外的层次语法信息，还可以对具有任意粒度的句子进行编码。

为了用树形长短期记忆模型表示依存树，我们首先用零初始化隐藏状态 h_j 和存储单元状态 c_j 。叶节点的输入向量用它们在预训练语言模型内的相应表示进行初始化，而非叶节点被设置为零。然后，我们使用树形长短期记忆模型 [59] 将叶节点的特征一直传播到句子的根节点。依存树中任意节点 j 的输入门 i_j 、输出门 o_j

① 依存语法树和依赖语法树通过 Stanza 库获得，这是目前最流行的依存语法树和依赖语法树库之一 <https://stanfordnlp.github.io/stanza/>

和遗忘门 f_{jk} 计算如下：

$$i_j = \sigma(W_i x_j + \sum_{l \in \text{Child}(j)} h_l W_{il} + b_i) \quad (3-2)$$

$$o_j = \sigma(W_o x_j + \sum_{l \in \text{Child}(j)} h_l W_{ol} + b_o) \quad (3-3)$$

$$f_{jk} = \sigma(W_f x_j + \sum_{l \in \text{Child}(k)} h_{jl} W_{kl} + b_f) \quad (3-4)$$

其中 σ 是 sigmoid 函数， W, b 是可训练的参数。 x_j 表示节点 j 的输入向量， $\text{Child}(j)$ 表示依存树中节点 j 的子节点。

每个节点的综合结果 u_j 计算如下：

$$u_j = \tanh(W_u x_j + \sum_{l \in \text{Child}(j)} h_l W_{ul} + b_u) \quad (3-5)$$

最后，我们更新隐藏状态 h_j 和存储单元状态 c_j 如下，

$$c_j = i_j \odot u_j + \sum_{l \in \text{Child}(j)} f_{jl} \odot c_l \quad (3-6)$$

$$h_j = o_j \odot \tanh(c_j) \quad (3-7)$$

在实践中，我们递归地更新每个节点的隐藏状态和存储单元状态，从根节点到叶节点。我们使用句子的根的特征向量作为整个句子的特征向量。对于具有 I 个句子的文档，我们可以获得所有的句子向量 $\{s_l\}_{l=1}^I$ 。由于并非所有的句子都包含用于关系推理的相关信息，我们在句子向量上使用多头注意力层来识别最相关的句子。我们将此过程表述为：

$$V_{doc} = [s_1, s_2, \dots, s_I] \quad (3-8)$$

$$S, A = \text{Attn}(W_{t1}(e_s - e_o), W_{t2}V_{doc}, W_{t3}V_{doc}) \quad (3-9)$$

其中 $W_{t1}W_{t2}$ 和 W_{t3} 是可训练的权重， Attn 表示注意力层， A 是注意力得分， S

是通过注意力层的句子向量。然后，我们将加权句子与实体对相结合，计算关系 r 的得分 z_{const} ：

$$z_s = \tanh(W_{(s1)}e_s + W_{(s2)}S) \quad (3-10)$$

$$z_o = \tanh(W_{(o1)}e_o + W_{(o2)}S) \quad (3-11)$$

$$z_{const} = z_s^\top W_{s,o} z_o + b_{const} \quad (3-12)$$

其中 Wb 是可训练的参数。在这里，我们获得了基于关注度得分 a 和句子向量 S 的依存树的关系得分 z_{const} ，以增强依赖图。

3.3 基于依赖语法树的文档级关系抽取

在获得上述表示后，我们构建了一个依赖语法图来聚合语法关联词的信息。文档中的每个句子都被输入到依赖语法分析器中，该分析器生成依赖语法树，我们同样利用 Stanza 库来完成这一语法分析器的识别。然后我们将依赖语法树转换为依赖语法图。在这一部分，我们添加了一个文档节点，以缩短实体对之间的距离，更好地聚合长距离句子的信息。

我们的依赖语法图包含四种节点：非根单词节点、根单词节点，实体提及节点和文档节点。具体地，文档中的每个单词对应于单词节点，并且对于不是依赖语法树中的句子根的单词，其编码特征对应于其节点特征。对于句子词根，我们特别利用公式3.2中获得的词根特征 S 来对抗可能的错误解析。我们使用多层感知器（MLP）将在公式3.2中原始特征与方程中的 S 融合，这是因为依赖语法图中的句子根应该包括整个句子的信息。对于实体提及的节点，通过对该实体提及中的单词的特征进行平均来计算节点特征。文档节点在这里作为包括每个句子词根的信息并强调有利句子的节点，被计算为每个句子向量的加权平均值，权重是公式3.2中实体对的注意力得分。

依赖语法图中有四种类型的边。其中三种边是双向边，一个是单向边。第一种边是双向依赖边，这种边在语法树中的每对连接的单词之间，我们在这对单词之间添加双向依赖性语法边。然后，第二种边是相邻句子之间的边。由于相邻句

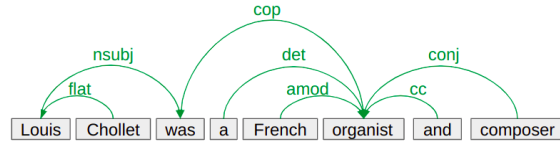


图 3-3 依赖语法树描述单个句子中单词之间的依赖语法关系。利用这样的语法信息可以显著地补充原始的纯文本，并捕捉相邻单词之间的关系。我们的模型将依赖语法树转换为依赖语法图

子之间存在密切的上下文关系，我们在相邻句子的依赖语法树根节点之间添加双向边。第三种边存在与句子根节点和文档节点之间。由于文档中的每个句子都服务于该文档的同一主题，因此在依赖语法树根根结和文档节点之间添加了双向边。最后一种边缘是单向的，存在于不相邻的句子根节点之间，用于捕获长距离信息和嵌入序列信息。

双向边的权重为 1，以表示它们的强连接。而非相邻句子根节点 i 到 j 的权重 $ADJ_{i,j}$ 由它们的特征向量相似度计算出来：

$$ADJ_{i,j} = \frac{S_i \cdot S_j}{\|S_i\| \cdot \|S_j\|} \quad (3-13)$$

其中 S_i 和 S_j 是公式 3.2 中的句子向量 i 和 j 。使用这些学习到的权重，我们的模型可以自动获得从文档前面句子的根节点到文档后面句子的根节点的逻辑连接，并融合依赖性和依存语法之间的信息，从而以不同的角度聚合信息。

在获得依赖图的相邻矩阵 ADJ 后，我们使用卷积图神经网络来进行特征聚合和实体增强。

$$q^{l+1} = GCN(q^l, ADJ) \quad (3-14)$$

其中 q^l 是层 l 的输入特征， q^{l+1} 是输出特征。

然后通过基于 $\log\text{sumexp}$ [22] 合并该实体的所有提及的嵌入来获得实体表示 $entity_i$ ：

$$entity_i = \log \sum_j \exp(m_{i,j}) \quad (3-15)$$

其中 $m_{i,j}$ 是 j^{th} 提及 $entity_i$ 的嵌入。

除了特定实体的嵌入之外，我们还提取了两个目标实体之间的最短路径 $path_{s,o}$ ，以补充依赖图上的实体对 (e_s, e_o) 。

$$path_{s,o} = [e_s, node_1, node_2, \dots, node_n, e_o] \quad (3-16)$$

为了提高效率，我们将所选路径的最大长度限制为 14，这意味着除了头部和尾部实体之外，最多将选择 12 个路径节点。任何长于 12 的路径都将选择前 12 个节点，而小于 12 的路径将用零张量填充。此外，我们还使用多层感知器来探索实体对之间的关系。

$$pair_{s,o} = LeakyRelu(W_{p1}e_s + W_{p2}e_o) \quad (3-17)$$

通过以上步骤，可以获得通过依赖语法补全的实体表示和上下文表示。按照过往的方法 [60]，我们将它们全部连接起来，以增强这个实体对的特征：

$$I_{s,o} = [e_s; e_o; pair_{s,o}; path_{s,o}] \quad (3-18)$$

我们基于依赖图计算关系得分 z_{dep} ：

$$z_{dep} = W_{(d2)}\sigma(W_{(d1)}I_{s,o} + b_{(d1)}) + b_{(d2)} \quad (3-19)$$

其中 W, b 是可训练的参数。

3.4 融合两种不同语法树结果的文档级关系抽取

大多数现有的关系提取模型输出 $P(R_i|e_s, e_o, D)$ 的概率作为实体对 (e_s, e_o) 对于关系 R_i 的预测值，这需要预先确定的全局阈值来将概率转换为关系标签。一些方法使用启发式阈值或学习在验证集上具有最高 F1 分数的全局阈值。然而，全局阈值可能不是对所有情况都是最优的，并且会引入误差。为了解决这个问题，Zhou 等人 [33] 引入了一个额外的阈值类作为阈值来区分正类和负类。然而，这种基于概率分布的方法可能会因为关系标签的长尾分布受到影响，其中大多数标签仅与少量训练示例相关联。即使预测是正确的（高于阈值），基于概率分布的损失（如交叉熵损失）仍可能引起大的损失。这是因为具有更多训练示例的主导类将对整个模型产生更强的影响，从而导致过度拟合主导类。

Hinge 损失 [61] 的许多变体已被提出，以克服各个领域多类学习任务中的长尾问题。Hinge 损失会训练一个最大裕度分类器，而不是对概率分布进行建模。

为了使得 Hinge 损失能够应用于我们的多标签分类任务，我们跟随 [14] 的思路，通过定义一种全新的标签来解决这个问题。给定一对实体 (e_s, e_o) ，我们首先将它们的关系标签划分为正类 P 和负类 N 。正类 P 包含存在于两个实体之间的关系。当这两个实体之间没有关系时，正类集合 P 可以是空的。负类集合 N 包含两个实体之间不存在的关系。在这里一组新标签被定义为

$$t_i = \begin{cases} 1, & R_i \in P \\ -1, & R_i \in N \end{cases} \quad (3-20)$$

在获得这个标签之后，我们将依赖语法图和依存语法树获得的两个分数组合为它们的动态加权和 [62]，权重 η 是可以自动调节获得的。

$$z_{final} = z_{dep} + \eta z_{const} \quad (3-21)$$

然后，我们将获得的新标签应用于损失函数中 [14]。

$$\mathcal{L} = \sum_{1 \leq i \leq C} \max(0, \alpha - t_i(z_{final}^i - z_{final}^s)) \quad (3-22)$$

其中 $\alpha > 0$ 是损失函数的超参数，来控制模型的过渡拟合， C 是类的数量。 z_{final}^i, z_{final}^s 分别是 $non\mathcal{NA}, \mathcal{NA}$ 类的预测结果。在这种情况下，该模型旨在做出“足够好”的预测，而不是“完美”的预测。因此，该模型避免了过度拟合任何类，尤其是对占主导地位类。

3.5 本章小结

在这一部分，我们介绍了我们是如何融合依赖语法和依存语法来改进文档级关系抽取。我们观察到，依赖语法更适合构建实体对之间的路径，而依存语法更适合聚合句子级信息。因此，我们遵循先前的研究，将依赖语法树转换为依赖语法图，并提取实体对之间的路径。然而，由于预训练模型中学习的语法与依赖树的黄金语法之间的差距严重影响了文档级关系抽取的性能，我们建议利用依存树来聚合句子级别的信息，以弥补依赖树中的差距。具体来说，我们使用单层感知器来将依存树中的句子根信息插入到依赖语法图中，以替换依赖语法图中的原始句子根。此外，为了更好地考虑依赖图中的句子交互，我们添加了一个文档级节点并链接每个句子根，以减少实体对的距离，更好地捕捉远距离关系。

第四章 基于语法树的文档级关系抽取实验结果

为了全面的评估我们的模型，我们在来自不同领域的三个文档级数据集上评估了我们的算法模型。我们在第 4.1 节介绍了这些数据集，在第 4.2 中介绍了我们的实现细节，在第 4.3 和 4.4 节说明了我们的结果。然后，我们在第 4.5 节分析了两种不同的语法树权重，在第 4.7 节分析了依赖语法树路径距离。最后，我们在第 4.6 节进行了消融实验，并总结了我们的研究成果。

4.1 文档级关系抽取数据集说明

统计数据	DocRED	CDR	GDA
# 训练集	3053	500	23353
# 验证集	1000	500	5839
# 测试集	1000	500	1000
# 关系数量	96	2	2
# 平均每篇文章的句子数量	8.0	9.7	10.2

表 4-1 三个公共数据集的统计数据

尽管文档级关系抽取方法可能因为要提取的关系类型（基因-疾病关系，一般类型的关系等）而有所不同，但大多数关系抽取方法都是在通用数据集和医学数据集上进行评估的。因此我们选择了一个通用数据集和两个医学数据集来评估我们的模型，表 4.1 列出了这些数据集的统计数据。

- **DocRED** [18] 是一个从维基百科中人为标注的大型数据集。它包含从维基百科中采样的 5053 个黄金注释文档，132275 个实体、56354 个关系事实和 96 个关系类，每篇平均长度为 196.7 个单词，超过 40.7% 的关系对是跨句关系事实。实体部分包括经典的实体标签，如 PERSON、ORGANIZATION 和 LOCATION；而实体关系部分使用了 96 种关系类型，涵盖了广泛的科目，包括科学（33.3%）、艺术（11.5%）、时间（8.3%）、个人生活（4.2%）等。
- **CDR** [19] 是一个生物医学的文档级关系抽取数据集，由 1500 篇 PubMed 的摘要组成。这些摘要被随机分为三个相等的部分进行训练、验证和测试。这个数据集的预测任务是预测化学品和疾病之间的二元关系。
- **GDA** [20] 也是一个生物医学的文档级关系抽取数据集，包含 30192 篇摘要。这个大型数据集是通过对公开数据库收集的基因-疾病关联中的 30192 篇摘要（测试集为 1000 篇，其余按 80:20 的比例在训练集和验证集之间划分）进

行远程监督构建的。这个数据集的预测任务是预测基因和疾病之间的二元关系。

4.2 DocRED 数据集实验结果说明与分析

我们的模型是在 Pytorch [63] 上实现的，并使用 stanza [64] 来提取依存和依赖语法。对于所有实验，我们的学习率设置为 $5e-5$ ，权重衰减为 $1e-4$ 。依赖关系图的卷积图神经网络的层数被设置为 3，并且输出维度为 128。依存树中每个节点的隐藏状态和单元状态共享 256 的维度。整个模型由 AdamW 优化。

4.3 DocRED 数据集实验结果说明与分析

我们将我们的模型与 DocRED 数据集中基于图的方法和基于注意力的方法进行了比较。对于基于注意力的方法，我们将所提出的方法与 BERT [18]、ATLOP [9]、EIDER [49] 和自训练方法 DREEAM [38] 进行了比较。基于图的模型包括 LSR [55]、HeterGSAN [21]、DRE [65]、CorefDRE [66]、GAIN [10]、SagDRE [14] 和 LARSON [12]。我们在 $BERT_{base}$ 和 $DeBERTa_{Large}$ 上训练我们的模型。我们不仅报告了 F1 和 Ign-F1 分数（不包括训练集和测试集共享的关系示例的 F1 分数），还报告了 Intra-F1（仅考虑句内关系的 F1 分数）和 Inter-F1（仅考虑句子间关系的 F1 分数）。

表4.3中列出的实验结果表明，我们的模型可以在一般领域的文档级关系抽取数据中取得领先的性能。在 Ign F1 和 F1 分数方面，所提出的模型在测试集上分别以 2.29% 和 1.94% 的优势优于基于依赖图的方法 SagDRE [14]，这表明依赖语法和依存语法的组合对于文档级关系提取是有用的。我们的模型将测试集上的 IgnF1 分数提高了 1.56%，而最先进的方法 LARSON [12] 使用依存树来预测 DocRE 的证据，并分别使用依存和依赖语法。这一进展证实了不同粒度的信息可以帮助 DocRE 中的关系提取。

4.4 医学数据集实验结果说明与分析

除了一般的领域文档级关系抽取方法外，我们还在两个生物医学领域数据集 CDR 和 GDA 上将我们的模型与各种先进的方法进行了比较，包括 LSR [55]、sciBERT [9]、ATLOP [9]、EIDER [49]、SagDRE [14] 和 LARSON [12]。

实验结果列于表4.4中。总之，我们的模型在两个测试数据集上实现了显著的改进（CDR 上为 0.93%，GDA 上为 1.27%）。GDA 和 CDR 的句子比 DocRED 多，因此我们的模型可以处理复杂的文档，进一步证明了它的通用性。

Model	dev				test	
	IgnF1	F1	Intra F1	Inter F1	IgnF1	F1
BERT _{base} [18]	-	54.16	61.61	47.15	-	53.20
LSR – BERT _{base} [55]	52.43	59.00	65.26	52.05	56.97	59.05
GAIN – BERT _{base} [10]	59.14	61.22	67.10	53.90	59.00	61.24
HeterGSAN – BERT _{base} [51]	58.13	60.18	-	-	57.12	59.45
DRN – BERT _{base} [65]	59.33	61.09	-	-	59.15	61.37
ATLOP – BERT _{base} [9]	59.22	61.09	-	-	59.31	61.30
CorefDRE – BERT _{base} [66]	60.85	63.06	-	-	60.78	60.82
EIDER – BERT _{base} [49]	60.51	62.48	68.47	55.21	60.42	62.47
SagDRE – BERT _{base} [14]	60.32	62.06	-	-	60.11	62.32
LARSON – BERT _{base} [12]	61.05	63.01	68.63	55.75	60.71	62.83
DREEAM – BERT _{base} [38]	60.51	62.55	-	-	60.03	62.49
FCDS-BERT_{base}	62.61	64.42	68.79	57.24	62.08	64.21
ATLOP – DeBERTa _{Large} [9]	62.16	64.01	68.45	59.63	62.12	64.08
FCDS-DeBERTa_{Large}	64.12	66.17	70.19	58.73	64.03	65.86

表 4-2 在 DocRED 数据集的验证集和测试集上关系预测的结果

Model	CDR	GDA
LSR-BERT [55]	64.80	82.20
SciBERT [9]	65.10	82.50
ATLOP-SciBERT [9]	69.40	83.90
EIDER-SciBERT [49]	70.63	84.54
SagDRE-SciBERT [14]	71.80	-
LARSON-SciBERT [12]	71.59	86.02
DREEAM-SciBERT [38]	71.55	84.51
FCDS-SciBERT	72.62	87.39

表 4-3 CDR 和 GDA 测试集上关系提取的结果

Model \ Metric	IgnF1	F1
LSR [55]	0.0146	0.0129
ATLOP [9]	0.0240	0.0213
EIDER [49]	0.0337	0.0313
LARSON [12]	0.0470	0.0439
DREEAM [38]	0.0352	0.0341

表 4-4 在所有数据集上进行两样本 t 检验。在 DocRED 数据集上使用 BERT 进行单词嵌入，在 CDR 和 GDA 上使用 SciBERT 进行单词嵌入

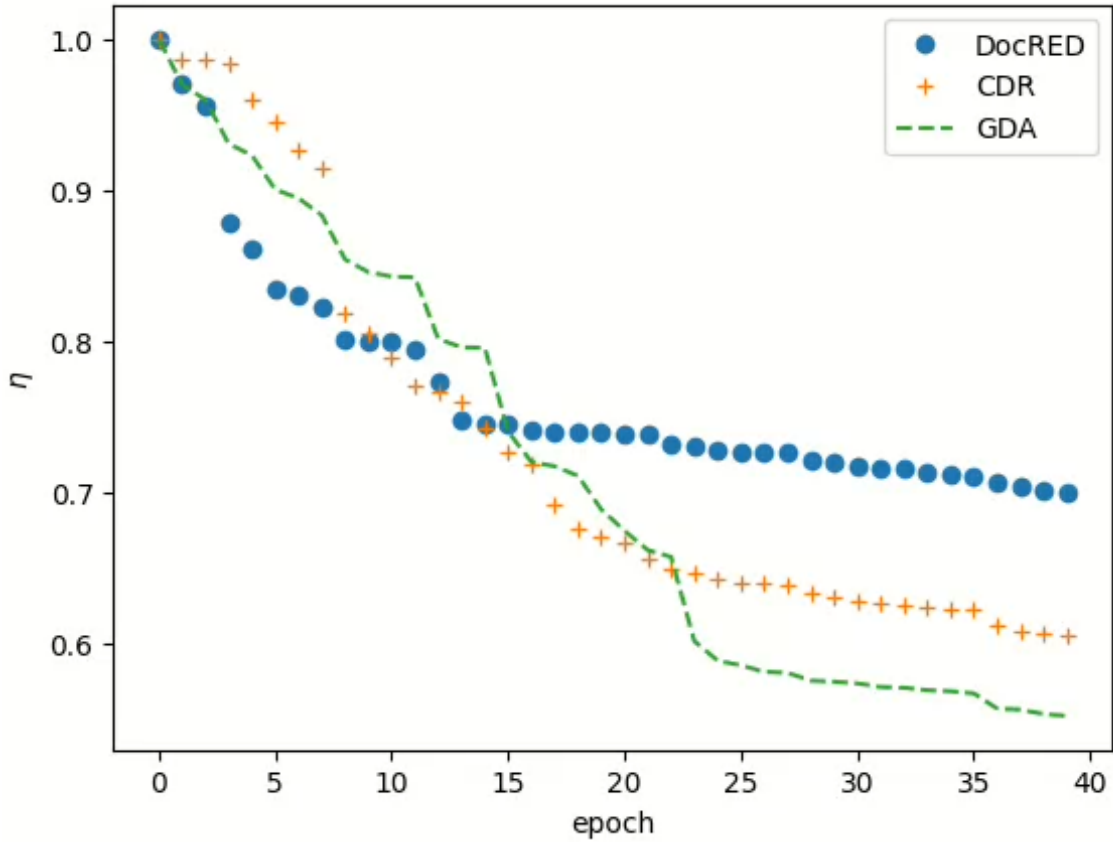


图 4-1 η 在 DocRED、CDR 和 GDA 数据集上的学习曲线

此外，为了评估改进的显著性，我们在三个数据集中进行了两样本 t 检验，将我们的方法与其他五种方法进行了比较。所获得的 p 值如表4.4所示。可以看出，所有值都小于 0.05，表明我们的方法有显著改善。

4.5 两种不同的语法树权重分析

为了全面了解我们方法的每个组成部分是如何对最终性能做出贡献的，我们进行了消融实验，以分析不同句法的功能。在图4.5中，我们绘制了方程中 η 的学

Ablation	F1	IgnF1
原方法	64.42	62.61
不对两种语法进行融合	62.59	60.63
取消文档级节点	62.78	60.92
取消依存语法树	62.15	60.22
取消依赖语法树	61.63	59.41

表 4-5 在 DocRED 的验证集中对我们的模型进行消融实验

	DocRED		CDR		GDA	
	具有文档节点	不具有文档节点	具有文档节点	w/o 不具有文档节点	具有文档节点	不具有文档节点
平均长度 & 标准差 d	6.19 ± 1.02	7.23 ± 1.55	5.96 ± 2.26	6.44 ± 2.23	6.82 ± 1.99	7.38 ± 2.43
最大长度	7	9	8	12	11	14
最小长度	4	4	4	4	3	4

表 4-6 DocRED、CDR、GDA 数据集中具有文档节点和不具有文档节点的实体距离的平均/最大/最小长度和标准差

习曲线。我们可以观察到 η 从初始值 1.0 开始减少，这表明依赖图在这个过程中发挥着更重要的作用，这可能是因为依赖图结合了依存语法树的特征并捕获了细粒度信息。

4.6 消融实验结果说明与分析

我们一次移除我们模型中的一部分，并使用表4.6中 DocRED 数据集中的验证集评估结果模型。对于单不对两种语法进行融合的情况，我们没有使用依存语法树来增强依赖语法图，并观察到 F1 和 IgnF1 分数分别减少了 0.78% 和 0.94%，这表明融合语法对文档级关系抽取有利。

然后，我们分别从文档级节点、依赖语法图和依存语法树之中删除一个。对于删除文档节点，我们观察到 F1 和 IgnF1 分别减少了 0.59% 和 0.65%。我们推测文档节点减少了实体对之间的距离，这有助于文档级关系抽取。对于删除依赖图，只有依存语法树被合并用于关系预测。我们可以观察到 F1 分数下降了 2.16%。当我们移除依存语法树时，也会出现类似的趋势，其中 F1 下降到 62.15%。因此，依赖语法图和依存语法树对我们的模型都很重要。

4.7 依赖语法树路径距离分析

为了检验文档节点对减少实体对之间距离的影响，我们分析了在表4.7中的三个数据集中添加文档级节点前后实体距离的变化。通过随机选择 600 个案例，我们评估了有和没有文档节点的实体距离的平均值、最大值、最小值和标准差。我们的发现表明，平均实体距离减少了，这表明文档节点可以通过简化实体之间的

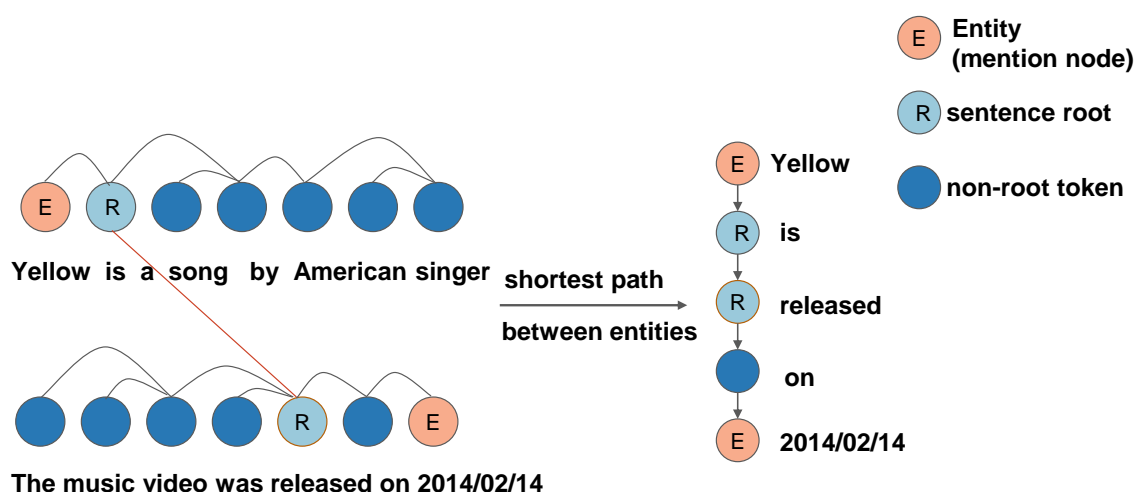


图 4-2 在这个例子中，依赖语法图通过提取实体 **Yellow** 和 **2014/02/14** 之间的最短路径：“**Yellow** is released on **2014/02/14**”，在文档级关系抽取中发挥了积极作用

交互来改进文档级关系抽取。

4.8 案例研究

为了更好地理解算法并启发未来的工作，我们进行了一个案例研究来研究我们算法预测的瓶颈。

图4.8中展示第一个例子是一个成功的依赖语法图，它说明了依赖语法如何帮助模型完成文档级关系抽取。在这种情况下，依赖语法分析器成功地解析了两句话中的依赖语法，并提取了重要的关键字来帮助预测。通过提取短路径“Yellow is released on 2014/02/14.”，模型只需要一个简单的句子级模型可以完成预测。

图4.8中展示的第二个例子说明了我们的方法背后的动机，这是为了减轻依赖分析器的不准确和为实体选择关键字失败的情况。尽管依赖语法对文档级关系抽取来说是一个有用的工具，但它有时无法识别实体的相关关键字。在这种情况下，最短路径“Yellow is was Chppa Boi”远未完成预测，而路径中未选择真正的关键字“producted”。此外，依赖语法解析器中的错误会导致信息的严重缺乏。为了解决这个问题，我们利用依存语法树并融合来自依赖语法图和依存语法的信息，它分层地组织单个句子的不同单词，并可以将句子级别的信息自然地聚合到句子根节点。通过这样做，句子根节点可以获得根节点本身和句子级信息，同时增加路

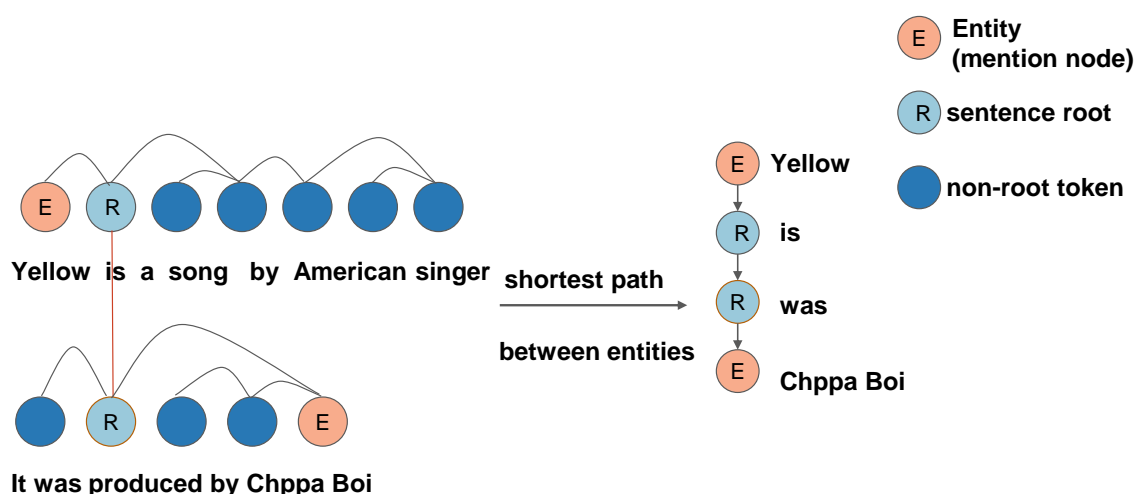


图 4-3 在这个例子中，依赖语法图在文档级关系抽取中扮演着矛盾的角色。它提取的实体对 **Yellow** 和其作者 **Chppa Boi** 之间的最短路径为：“**Yellow is was Chppa Boi**”。然而，由于依赖语法解析器的不准确，模型无法捕获到关键字 **produced**，因此很难预测实体之间的真实关系

径的信息，减轻依赖解析器的不准确度。

4.9 本章小结

在本章中，我们在三个不同的数据集上评估了我们提出的算法。我们发现，我们的模型在三个测试数据集上实现了显著的改进，在三个数据集中进行了两样本 t 检验，表明我们的方法有显著改善。我们还分析了不同语法的功能，并发现文档节点对实体距离的影响。最后，我们进行了一个案例研究来研究我们算法预测的瓶颈。

第五章 结论

在这项工作中，我们提出了一种新的文档级关系提取模型。我们的模型利用了两种类型的额外语法信息，即依赖语法和依存语法。我们采用图卷积网络和树形长短期记忆模型对这两类信息进行编码。此外，通过使用依存语法树来增强依赖语法图，并在依赖语法图中添加文档节点，可以提高依赖语法图的表达能力，从而更好地捕捉长距离相关性。在三个公共文档级关系提取数据集上的实验表明，我们的模型优于现有方法。

由于我们的研究重点是语法对文档级关系提取的影响，可能的未来的工作应该进一步分析依赖语法图和依存语法树等不同语法信息的影响。我们的模型在很大程度上依赖语法图，它提供了一种图形结构来连接语法相关的单词，并缩短两个实体之间的距离，但为什么依赖语法更有用，更受大多数研究人员的青睐，还有待进一步研究。另一方面，依存语法树是长期以来缺乏足够研究的语法。我们使用这些树来选择有利的句子，但还需要探索更有用的方法。

此外，提取重要句子和融合语法信息的影响不仅存在于文档级关系提取中，也存在于许多其他文档级任务（如阅读理解和文档检索）中。可能的未来工作可以进一步研究这些文档级任务的关联性。最后，提高基于图的方法的效率也会带来一些新的可能，这是因为目前大多数的图方法虽然取得了不错的结果，但是大多数图方法都是计算密集型的方法。

致 谢

在攻读计算机学士学位期间，首先热烈感谢我的导师康昭教授。经过风风雨雨的研究，我得到了他的无私关怀和支持。在此特别表达感谢之情。我还要感谢我的一直以来的帮助者们，包括学院的老师、同学、同事，以及所有支持和关心我的人。我也要感谢我的家人，他们给予我强大的内心支持和生活的安定和稳定。

参考文献

- [1] Dixit K, Al-Onaizan Y. Span-level model for relation extraction[C]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019, 5308-5314.
- [2] Lyu S, Chen H. Relation classification with entity type restriction[C]. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Online, 2021, 390-395.
- [3] Delaunay J, Tran T H H, González-Gallardo C-E, et al. A comprehensive survey of document-level relation extraction (2016-2022)[J]. arXiv preprint arXiv:2309.16396, 2023.
- [4] Han X, Wang L. A novel document-level relation extraction method based on bert and entity information[J]. IEEE Access, 2020, 8(): 96912-96919.
- [5] Huang Q, Zhu S, Feng Y, et al. Three sentences are all you need: Local path enhanced document relation extraction[C]. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Online, 2021, 998-1004.
- [6] Bai J, Wang Y, Chen Y, et al. Syntax-BERT: Improving pre-trained transformers with syntax trees[C]. Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Online, 2021, 3011-3020.
- [7] Gupta P, Rajaram S, Schütze H, et al. Neural relation extraction within and across sentence boundaries[C]. Proceedings of the AAAI conference on artificial intelligence, 2019, 6513-6520.
- [8] Ye D, Lin Y, Du J, et al. Coreferential Reasoning Learning for Language Representation[C]. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 2020, 7170-7186.
- [9] Zhou W, Huang K, Ma T, et al. Document-level relation extraction with adaptive thresholding and localized context pooling[C]. Proceedings of the AAAI conference on artificial intelligence, 2021, 14612-14620.
- [10] Zeng S, Xu R, Chang B, et al. Double graph based reasoning for document-level relation extraction[C]. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 2020, 1630-1640.
- [11] Liu H, Kang Z, Zhang L, et al. Document-level relation extraction with cross-sentence reasoning graph[C]. Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2023, 316-328.

- [12] Duan Z, Li X, Li Z, et al. Not just plain text! fuel document-level relation extraction with explicit syntax refinement and subsentence modeling[C]. Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, 2022, 1941-1951.
- [13] Sahu S K, Christopoulou F, Miwa M, et al. Inter-sentence relation extraction with document-level graph convolutional neural network[C]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019, 4309-4316.
- [14] Wei Y, Li Q. Sagdre: Sequence-aware graph-based document-level relation extraction with adaptive margin loss[C]. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2022, 2000-2008.
- [15] Sundararaman D, Subramanian V, Wang G, et al. Syntax-infused transformer and bert models for machine translation and natural language understanding[J]. arXiv preprint arXiv:1911.06156, 2019.
- [16] Xu Y, Jia R, Mou L, et al. Improved relation classification by deep recurrent neural networks with data augmentation[C]. Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 2016, 1461-1470.
- [17] Qin H, Tian Y, Song Y. Relation extraction with word graphs from n-grams[C]. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online and Punta Cana, Dominican Republic, 2021, 2860-2868.
- [18] Yao Y, Ye D, Li P, et al. DocRED: A large-scale document-level relation extraction dataset[C]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019, 764-777.
- [19] Li J, Sun Y, Johnson R J, et al. Biocreative v cdr task corpus: a resource for chemical disease relation extraction[J]. Database, 2016, 2016.
- [20] Wu Y, Luo R, Leung H C, et al. Renet: A deep learning approach for extracting gene-disease associations from literature[C]. Research in Computational Molecular Biology: 23rd Annual International Conference, RECOMB 2019, Washington, DC, USA, May 5-8, 2019, Proceedings 23, 2019, 272-284.
- [21] Zhang Z, Yu B, Shu X, et al. Document-level relation extraction with dual-tier heterogeneous graph[C]. Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain (Online), 2020, 1630-1641.

- [22] Jia R, Wong C, Poon H. Document-level n-ary relation extraction with multiscale representation learning[C]. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota, 2019, 3693-3704.
- [23] Gu J, Sun F, Qian L, et al. Chemical-induced disease relation extraction via convolutional neural network[J]. Database, 2017, 2017: bax024.
- [24] Li H, Yang M, Chen Q, et al. Chemical-induced disease extraction via recurrent piecewise convolutional neural networks[J]. BMC medical informatics and decision making, 2018, 18: 45-51.
- [25] Mandya A, Bollegala D, Coenen F, et al. Combining long short term memory and convolutional neural network for cross-sentence n-ary relation extraction[J]. arXiv preprint arXiv:1811.00845, 2018.
- [26] Su J, Wu Y, Ting H-F, et al. RENET2: high-performance full-text gene disease relation extraction with iterative training data expansion[J]. NAR Genomics and Bioinformatics, 2021, 3(3): 062.
- [27] Verga P, Strubell E, McCallum A. Simultaneously self-attending to all mentions for full-abstract biological relation extraction[C]. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), New Orleans, Louisiana, 2018, 872-884.
- [28] Wang H, Focke C, Sylvester R, et al. Fine-tune bert for docred with two-step process[J]. CoRR, 2019, abs/1909.11898.
- [29] Tang H, Cao Y, Zhang Z, et al. Hin: Hierarchical inference network for document-level relation extraction[J]. Advances in Knowledge Discovery and Data Mining, 2020, 12084: 197-209.
- [30] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data[C]. Advances in Neural Information Processing Systems, 2013, 2787-2795.
- [31] Kuang H, Chen H, Ma X, et al. A keyword detection and context filtering method for document level relation extraction[J]. Applied Sciences, 2022, 12(3): 2076-3417.
- [32] Yu J, Yang D, Tian S. Relation-specific attentions over entity mentions for enhanced document-level relation extraction[C]. Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, United States, 2022, 1523-1529.
- [33] Zeng D, Zhao C, Jiang C, et al. Document-level relation extraction with context guided mention integration and inter-pair reasoning[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022, 31: 3659-3666.

- [34] Xiao C, Yao Y, Xie R, et al. Denoising relation extraction from document-level distant supervision[C]. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 2020, 3683-3688.
- [35] Tan Q, He R, Bing L, et al. Document-level relation extraction with adaptive focal loss and knowledge distillation[C]. Findings of the Association for Computational Linguistics: ACL 2022, Dublin, Ireland, 2022, 1672-1681.
- [36] Xu B, Wang Q, Lyu Y, et al. Entity structure within and throughout: Modeling mention dependencies for document-level relation extraction[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(16): 14149-14157.
- [37] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. CoRR, 2017, abs/1706.03762.
- [38] Ma Y, Wang A, Okazaki N. Dream: Guiding attention with evidence for improving document-level relation extraction[C]. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Dubrovnik, Croatia, 2023, (toappear).
- [39] Quirk C, Poon H. Distant supervision for relation extraction beyond the sentence boundary[C]. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, Valencia, Spain, 2017, 1171-1182.
- [40] Peng N, Poon H, Quirk C, et al. Cross-sentence n-ary relation extraction with graph LSTMs[J]. Transactions of the Association for Computational Linguistics, 2017, 5: 101-115.
- [41] Tai K S, Socher R, Manning C D. Improved semantic representations from tree-structured long short-term memory networks[C]. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 2015, 1556-1566.
- [42] Christopoulou F, Miwa M, Ananiadou S. A walk-based model on entity graphs for relation extraction[C]. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Melbourne, Australia, 2018, 81-88.
- [43] Christopoulou F, Miwa M, Ananiadou S. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs[C]. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 2019, 4925-4936.
- [44] Guo Z, Zhang Y, Lu W. Attention guided graph convolutional networks for relation extraction[C]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019, 241-251.

- [45] Gupta P, Rajaram S, Schütze H, et al. Neural relation extraction within and across sentence boundaries[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(01): 6513-6520.
- [46] Tran H M, Nguyen M T, Nguyen T H. The dots have their values: Exploiting the node-edge connections in graph-based neural models for document-level relation extraction[C]. Findings of the Association for Computational Linguistics: EMNLP 2020, Online, 2020, 4561-4567.
- [47] Luan Y, Wadden D, He L, et al. A general framework for information extraction using dynamic span graphs[C]. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota, 2019, 3036-3046.
- [48] Wang H, Qin K, Lu G, et al. Document-level relation extraction using evidence reasoning on rst-graph[J]. Knowledge-Based Systems, 2021, 228: 107274.
- [49] Xie Y, Shen J, Li S, et al. Eider: Empowering document-level relation extraction with efficient evidence extraction and inference-stage fusion[C]. Findings of the Association for Computational Linguistics: ACL 2022, 2022, 257-268.
- [50] Li R, Zhong J, Xue Z, et al. Heterogenous affinity graph inference network for document-level relation extraction[J]. Knowledge-Based Systems, 2022, 250: 109146.
- [51] Xu W, Chen K, Zhao T. Document-level relation extraction with reconstruction[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 14167-14175.
- [52] Kim K, Hur Y, Kim G, et al. Greg: A global level relation extraction with knowledge graph embedding[J]. Applied Sciences, 2020, 10(3): 1181.
- [53] Li J, Xu K, Li F, et al. MRN: A locally and globally mention-based reasoning network for document-level relation extraction[C]. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Online, 2021, 1359-1370.
- [54] Zeng S, Wu Y, Chang B. SIRE: Separate intra- and inter-sentential reasoning for document-level relation extraction[C]. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Online, 2021, 524-534.
- [55] Nan G, Guo Z, Sekulic I, et al. Reasoning with latent structure refinement for document-level relation extraction[C]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 2020, 1546-1557.

- [56] Barreyro J P, Cevasco J, Burin D, et al. Working memory capacity and individual differences in the making of reinstatement and elaborative inferences[J]. The Spanish journal of psychology, 2012, 15(2): 471-479.
- [57] Zhang Y, Zhong V, Chen D, et al. Position-aware attention and supervised data improve slot filling[C]. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 2017, 35-45.
- [58] Kenton J D M-W C, Toutanova L K. Bert: Pre-training of deep bidirectional transformers for language understanding[C]. Proceedings of naacL-HLT, 2019, 2.
- [59] Miwa M, Bansal M. End-to-end relation extraction using LSTMs on sequences and tree structures[C]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 2016, 1105-1116.
- [60] Mou L, Men R, Li G, et al. Natural language inference by tree-based convolution and heuristic matching[C]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Berlin, Germany, 2016, 130-136.
- [61] Gentile C, Warmuth M K K. Linear hinge loss and average margin[C]. Advances in Neural Information Processing Systems, 1998, .
- [62] Kendall A, Gal Y, Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, 7482-7491.
- [63] Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. Advances in neural information processing systems, 2019, 32.
- [64] Qi P, Zhang Y, Zhang Y, et al. Stanza: A python natural language processing toolkit for many human languages[C]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Online, 2020, 101-108.
- [65] Xu W, Chen K, Zhao T. Discriminative reasoning for document-level relation extraction[C]. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Online, 2021, 1653-1663.
- [66] Xue Z, Li R, Dai Q, et al. Corefdre: Document-level relation extraction with coreference resolution[J]. arXiv preprint arXiv:2202.10744, 2022.