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。在本文中我们以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 effient 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 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。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，文档级关系提取识别文档中实体对之间的关系标签。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。

文档级关系提取面临的一个巨大障碍是推断长句中实体对的关系，这是因为长句中通常包含不相关甚至有噪声的信息。图 **??**是一个例子，它包括一对来自同一文档中的句子级关系和文档级关系。为了推断 Louis Chollet 和 Conservatoire de Paris 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 2 中的“He”一词指的是“Louis Cholet”。然而，由于现有的文档级关系提取模型 [3] 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。

最近，预训练语言模型在许多下游任务中显示出巨大的潜力。一些工作 [**?**,4] 试图通过预训练语言模型隐含地捕捉到实体之间的这种交互，但是最近的研究 [**?**,3,5] 表明通过预训练语言模型隐式学习的交互与黄金标记之间仍然存在很大差距。因此，对于文档级关系提取，隐含地学习一个有指导意义的上下文是不够的。而语法信息在句子级关系提取 [6,7] 中得到了广泛的应用，但在文档级关系提取的场景下尚未得到充分的探索。因此本文通过融合语法信息作为辅助，来改进文档级关系提取。

1.2 研究的主要内容

1.3 研究的主要贡献与创新

1.4 研究的结构安排

第二章 关系抽取研究

2.1 句子级关系抽取

2.2 文档级关系抽取

2.2.1 文档级关系抽取的预测目标

2.2.2 基于序列的文档级关系抽取

2.2.3 基于注意力机制的文档级关系抽取

2.2.4 基于语法树的文档级关系抽取

2.3 本章小结

1

第三章 基于语法树的文档级关系抽取研究

3.1 基于依存语法树的文档级关系抽取

3.2 基于依赖语法树的文档级关系抽取

3.3 融合两种不同语法树结果的文档级关系抽取

3.4 本章小结

1

第四章 基于语法树的文档级关系抽取实验结果

4.1 文档级关系抽取抽取数据集说明

4.2 DocRED 数据集实验结果说明与分析

4.3 医学数据集实验结果说明与分析

4.4 两种不同的语法树权重分析

4.5 依赖语法树路径距离分析

4.6 消融实验结果说明与分析

4.7 本章小结

1

第五章 结论

1

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] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.