

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，语法树融合

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 paper, 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, Syntax Tree Fusion

目 录

第一章 绪 论	1
第二章 关系抽取研究	2
第三章 语法信息研究	3
第四章 基于语法树的文档级关系抽取研究	4
第五章 基于语法树的文档级关系抽取实验结果	5
第六章 结论	6
致 谢	7
攻读本科学位期间取得的成果	9
外文资料原文	10
外文资料译文	11

第一章 绪 论

第二章 关系抽取研究

第三章 语法信息研究

第四章 基于语法树的文档级关系抽取研究

第五章 基于语法树的文档级关系抽取实验结果

第六章 结论

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。

ommand.

攻读本科学位期间取得的成果

- [1] Li J Y, **Zhang S**, Zhao Y W, et al. New memory method of impedance elements for marching-on-in-time solution of time-domain integral equation[J]. Electromagnetics, 2010, 30(5): 448-462.
- [2] 李四, **张三**. 时间步进算法中阻抗矩阵的高效存储新方法 [J]. 电波科学学报, 2010, 25(4): 624-631.
- [3] 李四, **张三**. 时域磁场积分方程时间步进算法稳定性研究 [J]. 物理学报, 2013, 62(9): 090206-1-090206-6.
- [4] **张三**, 李四. 时域磁场积分方程时间步进算法后时稳定性研究. 电子科技大学学报 [J] (已录用, 待刊).
- [5] **Zhang S**. Parameters discussion in two-level plane wave time-domain algorithm[C]. 2012 IEEE International Workshop on Electromagnetics, Chengdu, 2012, 38-39.
- [6] **张三**, 李四. 时域积分方程时间步进算法研究 [C]. 电子科技大学电子科学技术研究院第四届学术交流会, 成都, 2008, 164-168.
- [7] **张三** (4). 人工介质雷达罩技术研究. 国防科技进步二等奖, 2008 年.
- [8] XXX, XXX, XXX, XXX, 王升. XXX 的陶瓷研究. 四川省科技进步三等奖, 2003 年 12 月.

外文资料原文

1 The OFDM Model of Multiple Carrier Waves

外文资料译文

1 基于多载波索引键控的正交频分多路复用系统模型