****



**研 究 生 毕 业 论 文**

**（申请工程硕士学位）**

|  |  |
| --- | --- |
| **论文题目** | 基于深度学习的信息抽取服务的设  计与实现 |
| **作者姓名** | 安 磊 |
| **学科、专业名称** | 工程硕士(软件工程方向) |
| **研究方向** | 软件工程 |
| **指导教师** | 伏晓　副教授 |

**2019年 04 月 11 日**

**学 号： MF1732001**

**论文答辩日期： 年 月 日**

**指 导 教 师：**

**基于深度学习的信息抽取服务的**

**设计与实现**

|  |  |
| --- | --- |
| **作 者:** | **安磊** |
| 指导教师: | **伏晓　副教授** |

|  |
| --- |
| **南京大学研究生毕业论文** |
| **(申请工程硕士学位)** |

|  |
| --- |
| **南京大学软件学院** |
| **2019年04月** |

**The Design and Implementation of Information Extraction Service Based on Deep Learning**

**An, Lei**

**Submitted in partial fulfillment of the requirements for the degree of Master of Engineering**

Supervised by

Associate Professor **Fu, Xiao**

Software Institute

**NANJING UNIVERSITY**

Nanjing, China

Apr, 2019

# 摘 要

随着人工智能技术的快速发展，越来越多的算法得到应用。自然语言处理作为人工智能领域中的一个重要分支一直是许多人关注的焦点，而知识图谱是自然语言处理中一个重要的研究方向。构建知识图谱的数据来源包括结构化文本和非结构化文本，作为构建知识图谱的重要数据来源之一的非结构化文本，如何从中提取高质量的信息成为一个研究热点。

面对上文所述的问题，一个能够提供高质量数据的信息抽取服务就显得十分重要。本文以从非结构化文本中进行信息抽取为背景，以构建知识图谱时所需的三元组作为目标，实现从非结构化数据到知识图谱三元组的数据处理与转换。

本文基于构建非结构化文本的信息抽取服务的具体项目。从模型设计和具体的使用场景出发，本文选择了两种不同的命名实体识别模型和一种关系抽取模型进行组合，并将模型封装在Docker容器中，部署到Kubernetes集群上构建一个信息抽取服务系统。

本文首先介绍了信息抽取的两个子任务命名实体识别和关系抽取的研究背景，阐述了本文的一些主要工作。然后本文对实现信息抽取服务中运用到的一些技术进行简单的介绍与说明。之后根据信息抽取服务系统的架构进行需求分析，包括详细的需求分析、项目的架构分析以及模型设计的详细分析。此后介绍了每个模块的具体实现过程，对所有模型进行实验分析，信息抽取服务进行部署。最后对整个项目进行总结，并对未来的研究工作进行规划与展望。

**关键词**：信息抽取，命名实体识别，关系抽取，深度学习，服务

# Abstract

With the rapid development of artificial intelligence, more and more algorithm were applied.Natural Language Processing has always been the focus of many people as an important branch in the field of artificial intelligence,and Knowledge Graph is an important research direction in Natural Language Processing.The data sources for constructing Knowledge Graph include structured text and unstructured text.As an unstructured text that is one of the important data sources for constructing Knowledge Graphs,how to extract high quality information from it becomes a research hot spot.

In the face of the problems described above,an information extraction service that provides high quality data is very important.Based on the background of information extraction from unstructured texts,this thesis aims to construct data graphs form the unstructured data to the Knowledge Graph triples.

This thesis is based on the specific project of information extraction service for unstructured texts.From the model design and specific usage scenarios,the thesis combine two different Named Entity Recognition models and one Relation Extraction model,package the models in a Docker container and deploy to the Kubernetes cluster to build an information extraction service system.

The thesis firstly introduces the research background of information extraction`s two sub-task Named Entity Recognition and Relation Extraction,explain some main work of this thesis.Then,the thesis briefly introduces and explains some technologies used in the implementation of information extraction service.After that,the thesis describes the requirements analysis,including detailed requirements analysis,project architecture analysis and detailed analysis of model design.Next,the thesis introduces the specific implementation process of each module,conduct experimental analysis on all models and deploy information extraction service.Finally,the thesis summaries the entire project,plans and forecasts the future research work.

**Keywords:**Information Extraction, Named Entity Recognition, Relation Extraction, Deep Learning, Service

目录

[第一章 绪论](#_Toc4854_WPSOffice_Level1) [1](#_Toc4854_WPSOffice_Level1)

[1.1 项目背景](#_Toc443_WPSOffice_Level2) [1](#_Toc443_WPSOffice_Level2)

[1.2 国内外研究现状](#_Toc7644_WPSOffice_Level2) [2](#_Toc7644_WPSOffice_Level2)

[1.2.1 命名实体识别研究现状](#_Toc443_WPSOffice_Level3) [2](#_Toc443_WPSOffice_Level3)

[1.2.2 关系抽取研究现状](#_Toc7644_WPSOffice_Level3) [4](#_Toc7644_WPSOffice_Level3)

[1.3 本文主要的工作](#_Toc29847_WPSOffice_Level2) [6](#_Toc29847_WPSOffice_Level2)

[1.4 本文的组织结构](#_Toc14334_WPSOffice_Level2) [6](#_Toc14334_WPSOffice_Level2)

[第二章 技术综述](#_Toc443_WPSOffice_Level1) [8](#_Toc443_WPSOffice_Level1)

[2.1 引言](#_Toc31796_WPSOffice_Level2) [8](#_Toc31796_WPSOffice_Level2)

[2.2 词向量](#_Toc5318_WPSOffice_Level2) [8](#_Toc5318_WPSOffice_Level2)

[2.2.1 基于矩阵的分布表示](#_Toc29847_WPSOffice_Level3) [9](#_Toc29847_WPSOffice_Level3)

[2.2.2 基于神经网络的分布表示](#_Toc14334_WPSOffice_Level3) [9](#_Toc14334_WPSOffice_Level3)

[2.2.3 词向量的使用](#_Toc31796_WPSOffice_Level3) [10](#_Toc31796_WPSOffice_Level3)

[2.3 双向长短时记忆模型](#_Toc10788_WPSOffice_Level2) [11](#_Toc10788_WPSOffice_Level2)

[2.3.1 循环神经网络](#_Toc5318_WPSOffice_Level3) [11](#_Toc5318_WPSOffice_Level3)

[2.3.2 长短时记忆模型](#_Toc10788_WPSOffice_Level3) [12](#_Toc10788_WPSOffice_Level3)

[2.3.3 双向长短时记忆模型](#_Toc24764_WPSOffice_Level3) [13](#_Toc24764_WPSOffice_Level3)

[2.4 深度残差网络](#_Toc24764_WPSOffice_Level2) [13](#_Toc24764_WPSOffice_Level2)

[2.5 注意力机制](#_Toc23819_WPSOffice_Level2) [14](#_Toc23819_WPSOffice_Level2)

[2.5.1 Encoder-Decoder框架](#_Toc15178_WPSOffice_Level3) [14](#_Toc15178_WPSOffice_Level3)

[2.5.2 注意力模型](#_Toc19637_WPSOffice_Level3) [15](#_Toc19637_WPSOffice_Level3)

[2.6 远程监督](#_Toc15178_WPSOffice_Level2) [17](#_Toc15178_WPSOffice_Level2)

[2.7 深度学习框架Tensorflow](#_Toc19637_WPSOffice_Level2) [18](#_Toc19637_WPSOffice_Level2)

[2.9 本章小结](#_Toc25326_WPSOffice_Level2) [18](#_Toc25326_WPSOffice_Level2)

[第三章 基于深度学习的命名实体识别模型](#_Toc7644_WPSOffice_Level1) [19](#_Toc7644_WPSOffice_Level1)

[3.1 模型的整体架构](#_Toc5123_WPSOffice_Level2) [19](#_Toc5123_WPSOffice_Level2)

[3.2 模型的标注策略](#_Toc27641_WPSOffice_Level2) [20](#_Toc27641_WPSOffice_Level2)

[3.3 输入层](#_Toc3884_WPSOffice_Level2) [22](#_Toc3884_WPSOffice_Level2)

[3.3.1 字向量](#_Toc7744_WPSOffice_Level3) [22](#_Toc7744_WPSOffice_Level3)

[3.3.2 额外特征](#_Toc22826_WPSOffice_Level3) [23](#_Toc22826_WPSOffice_Level3)

[3.4 网络层](#_Toc7744_WPSOffice_Level2) [24](#_Toc7744_WPSOffice_Level2)

[3.5 输出层](#_Toc22826_WPSOffice_Level2) [25](#_Toc22826_WPSOffice_Level2)

[3.5.1 全连接层](#_Toc18750_WPSOffice_Level3) [25](#_Toc18750_WPSOffice_Level3)

[3.5.2 CRF层](#_Toc7396_WPSOffice_Level3) [25](#_Toc7396_WPSOffice_Level3)

[3.6 实验](#_Toc14704_WPSOffice_Level2) [27](#_Toc14704_WPSOffice_Level2)

[3.6.1 实验数据与采集](#_Toc6646_WPSOffice_Level3) [27](#_Toc6646_WPSOffice_Level3)

[3.6.2 实验评价指标](#_Toc30499_WPSOffice_Level3) [31](#_Toc30499_WPSOffice_Level3)

[3.6.3 模型训练的实现](#_Toc27939_WPSOffice_Level3) [31](#_Toc27939_WPSOffice_Level3)

[3.6.4 模型参数设置](#_Toc636_WPSOffice_Level3) [37](#_Toc636_WPSOffice_Level3)

[3.6.5 不同模型的性能对比](#_Toc15507_WPSOffice_Level3) [38](#_Toc15507_WPSOffice_Level3)

[3.7 本章小结](#_Toc18750_WPSOffice_Level2) [39](#_Toc18750_WPSOffice_Level2)

[第四章 基于深度学习的关系抽取模型](#_Toc29847_WPSOffice_Level1) [40](#_Toc29847_WPSOffice_Level1)

[4.1 模型的整体架构](#_Toc7396_WPSOffice_Level2) [40](#_Toc7396_WPSOffice_Level2)

[4.2 输入层](#_Toc6646_WPSOffice_Level2) [41](#_Toc6646_WPSOffice_Level2)

[4.2.1 字向量](#_Toc1426_WPSOffice_Level3) [41](#_Toc1426_WPSOffice_Level3)

[4.2.2 相对位置向量](#_Toc25146_WPSOffice_Level3) [41](#_Toc25146_WPSOffice_Level3)

[4.3 卷积层](#_Toc30499_WPSOffice_Level2) [42](#_Toc30499_WPSOffice_Level2)

[4.4 残差卷积块](#_Toc27939_WPSOffice_Level2) [42](#_Toc27939_WPSOffice_Level2)

[4.5 最大池化层](#_Toc636_WPSOffice_Level2) [44](#_Toc636_WPSOffice_Level2)

[4.6 输出层](#_Toc15507_WPSOffice_Level2) [44](#_Toc15507_WPSOffice_Level2)

[4.7 实验](#_Toc16217_WPSOffice_Level2) [44](#_Toc16217_WPSOffice_Level2)

[4.7.1 实验数据](#_Toc20460_WPSOffice_Level3) [44](#_Toc20460_WPSOffice_Level3)

[4.7.2 实验评价标准](#_Toc16453_WPSOffice_Level3) [46](#_Toc16453_WPSOffice_Level3)

[4.7.3 模型训练的实现](#_Toc27465_WPSOffice_Level3) [46](#_Toc27465_WPSOffice_Level3)

[4.7.4 模型参数设置](#_Toc11139_WPSOffice_Level3) [53](#_Toc11139_WPSOffice_Level3)

[4.7.1 不同模型的性能对比](#_Toc18104_WPSOffice_Level3) [55](#_Toc18104_WPSOffice_Level3)

[4.8 本章小结](#_Toc1426_WPSOffice_Level2) [55](#_Toc1426_WPSOffice_Level2)

[第五章 总结与展望](#_Toc14334_WPSOffice_Level1) [56](#_Toc14334_WPSOffice_Level1)

[5.1 总结](#_Toc25146_WPSOffice_Level2) [56](#_Toc25146_WPSOffice_Level2)

[5.2 进一步工作展望](#_Toc11644_WPSOffice_Level2) [56](#_Toc11644_WPSOffice_Level2)

[参 考 文 献](#_Toc31796_WPSOffice_Level1) [57](#_Toc31796_WPSOffice_Level1)

[致 谢](#_Toc5318_WPSOffice_Level1) [61](#_Toc5318_WPSOffice_Level1)

[版权及论文原创性说明](#_Toc24764_WPSOffice_Level1) [62](#_Toc24764_WPSOffice_Level1)

# 第一章 绪论

## 1.1 项目背景

自然语言处理（Nature Language Processing，简称NLP）是用计算机来处理、理解以及运用人类语言（如中文、英文等），它属于人工智能（Artificial Intelligence，简称AI）的一个发展方向，是计算机科学和语言学的交叉学科，又常被称为计算机语言学。由于自然语言是人类与其他动物的区别的主要标志，没有语言，人类的思维也就无从可谈，所以自然语言处理体现了人工智能的最高任务与境界。也就是说，只有当计算机具备了处理自然语言的能力时，机器才算实现了真正的智能。近十年，随着计算机设备和硬件的发展、计算设备的算力极大提升，自然语言处理的研究得到了很大的发展，并在各个领域都有了广泛应用。

信息抽取作为自然语言处理中的一项基础任务，在许多上层应用中都有着十分重要的影响。在上市公司新闻舆情分析中的关键信息就是新闻文本中的实体，如：公司、人、行业等实体信息以及实体间的关系信息，利用信息抽取中的命名实体识别和关系抽取就可以从新闻文本中抽出这些重要信息，进而对新闻舆情进行后续分析；在搜索引擎上进行搜索时的关键信息就是输入的搜索信息中的实体信息，但只通过分词进行输入文本的切分很容易把一个多词组成的命名实体切分成多个词，造成搜索不准确甚至搜索的结果完全不符合预期。如果能对输入文本进行命名实体识别就可以准确识别搜索的内容，再建立索引进行搜索，可以提高搜索的准确率也可以提升搜索的效率；在利用非结构化文本信息构建知识图谱的任务中关键信息就是非结构化文本中的关联实体以及实体间的关系组成的三元组，如：从“阿里巴巴集团的创始人是马云”这段描述中可以提取出“阿里巴巴，创始人，马云”这个三元组，可以表示“阿里巴巴的创始人是马云”这条信息。这里即用到了命名实体识别识别出了阿里巴巴和马云这两个实体，又用到了关系抽取来对阿里巴巴和马云这两个实体的关系进行识别和抽取。最后从海量非结构化文本中提取出大量三元组，经过实体消歧、知识融合后构建成一张知识图谱。

命名实体识别（Named Entity Recognition，简称NER）作为信息抽取中的一个基本任务，最早是在MUC-6（the Sixth of the Message Understanding Conference）上将其作为信息抽取的一项子任务引入测评中[MUC-6, 1996]。关系抽取（Relation Extraction，简称RE）作为信息抽取中的另一个基本任务，早在2000年就有基于句法解析增强的方法来实现关系抽取任务[Miller et al., 2000]。而本文同时结合命名实体识别和关系抽取两项任务作为利用非结构化文本中的三元组信息来构建知识图谱的关键环节，如何确保实体识别和关系抽取的准确性，提供高效稳定的信息抽取服务是构建高质量、高精度的知识图谱的关键所在。

## 1.2 国内外研究现状

## 1.2.1 命名实体识别研究现状

自从在MUC-6 [MUC-6, 1996]上将命名实体识别作为信息抽取的一个子任务之后，随着数据时代的到来，命名实体在各个领域的信息抽取任务中都得到了充分的利用和发展。

在国外命名实体识别的研究中，考虑到英文本身的特性，英文的命名实体识别只需要考虑词本身的特征而不需要关注别的问题例如分词等问题的影响，所以实现的难度相对较小，准确率也较高，

中文的命名实体识别由于中文文字内在的特殊性导致了在做命名实体识别之前要先做分词、句法分析等步骤，导致中文命名实体相对英文而言难度大大提高。在MUC-6中国内专家学者们把“人名”、“地名”、“组织机构名”作为实体的名称表达式，将这三类实体作为主要研究目标。在随后的研究中，逐渐地有人将“时间”、“货币”、“百分数”等也作为各种实体的表达式加入到命名实体识别的任务中来。

最早在命名实体识别任务中，主要有基于专家知识构建规则的方法以及基于数理统计的方法两类。基于词典和规则的方法主要是靠人工建立规则体系，在构建了大量基于专家知识的词法和语义规则之后，系统会根据规则对输入文本进行解析，基于人工构建的有限规则集合对文本中可能的命名实体进行推理和识别。虽然在特定的数据集上基于专家知识构建规则的方法相对于基于统计方法的命名实体识别而言准确度更高，但这种方法在数据量逐渐增大、数据内容逐渐复杂之后会变得不再可行，因为基于某一小部分语料构建的规则体系在别的语料上讲不再适用，人们无法去扩充和维护一个十分庞大的规则体系，且随着近些年机器学习理论的不断完善和计算机计算性能的提高，大部分学者开始转而研究基于统计方法的命名实体识别。

基于统计方法的命名实体识别主要有有监督（Supervised）、半监督（Semi-Supervised）的机器学习方法，其中有监督的机器学习方法在拥有大批量的标注语料的前提下在不同领域的文本中都拥有更高的识别准确率和精度，从而被更多的学者、从业者们所广泛使用。有监督学习方法中在命名实体识别任务中表现比较出色的方法主要有：基于隐马尔科夫模型（Hidden Markov Models，简称HMM）的命名实体识别模型[Bikel et al., 1999]、最大熵模型（Maximum Entropy Models，简称MEM）[Tsai et al., 2004]、条件随机场（Conditional Random Fields，简称CRFs）[McCallum et al., 2003]。基于统计方法的命名实体识别对特征选取的要求比较高，需要从文本中挖掘出对于实体识别有用的单词信息、上下文信息、句法信息、语义信息等作为特征。随着研究的不断深入，大量实验结果表明条件随机场结合了HMM和MEM的有点，成为中文命名实体识别任务中表现更优秀的统计学习方法。有学者针对条件随机场的特征选择与交叉组合进行了研究，通过实验得出不同的特征以及特征组合在训练时的贡献大小[张祝玉等, 2008]。

中文命名实体识别（Chinese Named Entity Recognition，简称CNER）相对于英文命名实体识别任务仍存在许多不同，由于中文本身的特点造成许多特殊的难点：

1. 中文单词的边界相比于英文较模糊

英文中的单词之间都有分隔符来标识边界，命名实体如人名、地名等单词的首字母为大写，这些信息能很好的为命名实体的识别做边界和位置标识，而中文单词没有这些信息。

1. 中文命名实体的词语结构更加复杂

有的类型的命名实体单词长度没有限制；不同实体有不同的组成结构（嵌套、简称、别名等）；音译词没有统一的构词规范；组织机构名和人名、地名会有很多交叉重复的地方（如：星环信息科技（上海）有限公司）。

1. 中文命名实体的一词多义与一义多词

随着互联网的极速发展，越来越多的网络词汇不断出现，很多原本的命名实体在现在都有了新的称呼（如称淘宝为某宝），这些词汇的新称呼的出现大大增加了命名实体识别的难度。

随着深度学习的影响不断增加，用深度学习方法解决命名实体识别任务也获得的一些显著的成果。[Wu et al., 2015]提出使用深度神经网络（Deep Neural Network，简称DNN）从语料中训练词向量，再输入到另一个深度神经网络中进行命名实体识别，实验结果好于传统机器学习中效果最好的CRF模型。[Z Huang et al., 2015]提出了使用双向长短时记忆模型(Bidirectional Long-Short Term Memory Model，简称Bi-LSTM)和条件随机场进行命名实体识别。还有一系列深度学习方法对命名实体识别任务做了很多尝试，在不同的数据集上相较于传统方法都有不小的提升。在NLP任务中，通常会先使用预训练好的词向量来进行后续任务，因为预训练好的词向量可以提取到词级别的特征，有助于提升后续任务的效果。Google在2018年提出了BERT模型[Devlin J et al., 2018]作为NLP任务中的预训练模型，结合对下游任务的微调在NLP的各项任务中的表现都达到了SOTA（state-of-the-art，最先进水平的）。表1.1为目前命名实体识别的一些常用方法汇总。

表 1.1 常用命名实体识别方法汇总

|  |
| --- |
| 模型名称 |
| CRF |
| DNN |
| LSTM |
| BiLSTM |
| LSTM+CRF |
| BiLSTM+CRF |
| BERT+BiLSTM+CRF |

## 1.2.2 关系抽取研究现状

关系抽取的方法主要有：

1. 有监督的学习方法

把关系抽取任务当做一个分类问题，从训练数据中设计并提取有效的特征并训练学习各种分类模型进行关系预测。需要大量的人工标注语料，比较耗时耗力，但预测效果较好。

1. 半监督的学习方法

先人工设定若干种子实例，再迭代地从文本中抽取相应的关系模板和更多关系实例。

1. 无监督的学习方法

假设有相同语义的命名实体同样拥有相似的上下文信息，可根据每个命名实体对的上下文信息来代表它们的语义关系，并对所有实体对的语义进行聚类。

与另外两种方法相比，有监督的学习方法在抽取的准确率和召回率上都有着更好的表现。

自从2000年Miller等提出基于句法解析增强的方法来实现关系抽取[Miller et al., 2000]后，越来越多的实体间关系抽取方法被提出：基于逻辑回归的方法[Kambhatla et al., 2004]、基于核函数的方法[Zhao and Grishman, 2005]、基于条件随机场的方法[Culotta et al., 2006]。在有监督学习中针对需要大量人工标注的情况下，Mintz等人[Mintz et al., 2009]提出了使用远程监督（Distant Supervision）的方法来扩充标注语料，可以有效解决关系抽取的标注数据规模问题。

随着计算机硬件和计算能力的极速发展，深度学习开始逐渐被学术界和业界所关注，很多学者也开始将深度学习运用到关系抽取任务中。[Socher et al., 2012]提出使用递归神经网络（Recurrent Neural Network，简称RNN）来解决关系抽取问题，通过递归神经网络学习到句子的词汇特征、句法特征、语义特征再用于关系分类和抽取。[Zeng et al., 2014]提出使用卷积神经网络（Convolutional Neural Network，简称CNN）来解决关系抽取问题，采用词向量和词的相对位置作为卷积神经网络的输入，通过卷积、池化、非线性计算等操作得到句子表示并用于关系抽取。[Santos et al., 2015]提出了一种新的卷积神经网络结构用于解决关系抽取问题，在这个新的结构中采用了新的Ranking损失函数并得到了更好的效果。[Miwa er al., 2016]提出一种基于端到端（End to End）神经网络的关系抽取模型，使用双向长短时记忆模型和树形长短时记忆模型同时对实体和句子进行建模。[Lin et al., 2016]提出基于句子级注意力机制的神经网络模型来解决关系抽取问题，文中的方法可以根据不同关系为每个实体对分配不同的权重。关系抽取的一些方法汇总如表1.2所示。

表1.2 关系抽取方法汇总

|  |
| --- |
| 模型名称 |
| RNN |
| CNN |
| CNN+Attention |
| BiLSTM+Attention |
| PCNN+Attention |
| ResCNN |

## 1.3 本文主要的工作

本文在总结信息抽取现有的一些研究成果的基础上，结合命名实体识别和关系抽取两项信息抽取的基本任务，从多个深度学习模型中选择合适的模型，利用人工标注的部分金融领域新闻文本语料结合远程监督进行模型训练，将训练好的模型固化成pb模型，把模型进行组合封装到Docker中，并部署在集群上来提供信息抽取服务。

本文的主要工作有：

1. 对命名实体识别和关系抽取的研究现状和相关技术进行综述，对命名实体识别模型和关系抽取模型进行实验对比分析。
2. 对本文的信息抽取服务系统进行了详细设计，包括需求分析、产品功能分析、本文选择的信息抽取模型的详细设计以及系统中每个模块的详细设计。
3. 利用远程监督扩充人工标注数据，缓解人工标注数据费时费力的缺点。
4. 将训练好的模型封装成ModelServer部署到集群上，构建一套信息抽取服务系统。

## 1.4 本文的组织结构

本文的各章节内容安排如下：

第一章 绪论。本章介绍了论文的项目背景，信息抽取任务中的两个子任务命名实体识别和关系抽取，国内外在这两个方向上的研究成果，本文所做的主要工作和论文的组织结构。

第二章 技术综述。本章详细介绍了论文涉及的相关技术，包括词向量、BERT模型、双向长短时记忆模型、深度残差网络、远程监督、深度学习框架Tensorflow、Docker容器技术、flask框架、Kubernetes集群等。

第三章 信息抽取服务分析与设计。本章详细介绍了信息抽取服务的总体规划、信息抽取服务的需求分析、信息抽取服务中各个模块的详细设计以及信息抽取模型结构的设计。

第四章 信息抽取服务的实现。本章阐述了信息抽取服务中各个模块的详细实现，包括数据预处理模块、信息抽取模型模块、信息抽取服务模块。

第五章 实验对比与分析。本章主要通过实验介绍了项目中选择的模型在不同参数下的模型效果对比以及不同模型之间的性能对比。

第六章 总结与展望。总结在论文期间做了哪些工作，在信息抽取任务中未来的可改进的地方以及对接下来的需要做的工作的进一步展望。

# 第二章 技术综述

## 2.1 引言

本章主要介绍了本文中实现的基于深度学习的命名实体识别和关系抽取所涉及的相关技术，包括词向量、双向长短时记忆模型、深度残差网络、注意力机制和远程监督。针对这些模型和算法的各自特点和实际使用过程，分析了它们的优缺点。

## 2.2 词向量

因为计算机不能直接识别自然语言，所以要想让计算机能够读懂自然语言，必须要对自然语言进行进一步处理，因此词向量（Word Embedding）的概念被提出。计算机能够识别的是二进制数，词向量就是一些包含了自然语言中词汇和语义信息的计算机能够识别的数字。

最早的词表示方法是One-Hot编码，即用N个状态编码器对N个状态进行编码，这种方法会造成一篇篇幅有n个字的文本，会用一个n\*n的矩阵来表示，每个n维向量中只有它对应的那个字的位置上是1，其余位置都是0。这种词表示方法在传统机器学习中，如条件随机场、逻辑回归、最大熵、支持向量机等模型中可以较好的完成一些自然语言处理的任务。但是这种词表示方法中默认每个单词之间是独立存在的，忽略了文本中单词间的语义关联，且通常向量的维度会很高，在用深度学习处理自然语言处理问题时如果使用One-Hot编码容易造成“维度灾难”[Bengio et al. 2002]。

除了忽略了单词间语义信息的One-Hot编码，[Hinton, 1986]提出了一种新的词表示方法：分布式表示法（Distributed Reputation）。基于[Harris, 1954]提出的分布假说：上下文相似的词，其语义也相似。词的分布式表示主要分为2类：基于矩阵的分布表示和基于神经网络的分布表示。

## 2.2.1 基于矩阵的分布表示

基于矩阵的分布表示主要是构建“词-上下文”矩阵，通过某种技术从矩阵中获取词的分布表示。其中，矩阵的每一行都表示一个词，列表示上下文信息。常见的上下文有：

1. 文档，即“词-文档”矩阵。
2. 上下文的每个词，即“词-词”矩阵。
3. n-元词组，即“词-n-元词组”矩阵。

矩阵中的每个元素表示词和上下文共现的次数，可以利用词频-逆文本频率指数（TF-IDF）、取对数等技巧进行加权和平滑。矩阵维度较高且较稀疏的情况还可以利用奇异值分解（SVD）和非负矩阵分解（NMF）等方法进行分解降维使之变成低维稠密矩阵。基于矩阵的分布表示的典型代表是Global Vector模型（GloVe模型）[Pennington et al. 2014]。

## 2.2.2 基于神经网络的分布表示

基于神经网络的分布表示一般通过神经网络训练语言模型得到的产物，[Bengio et al. 2006]提出了一种神经网络语言模型（Neural Network Language Model。简称NNLM），考虑对语言模型进行建模。但是NNLM中词向量知识语言模型训练得到的副产物，并没有指出哪一套向量作为词向量的效果更好，所以[Mikolov et al. 2013a; Mikolov et al. 2013b]在NNLM等模型的基础之上，提出了word2vec模型，在word2vec模型中设计了CBOW（Continuous Bag-of-Words）和Skip-gram两个模型，高效地训练获取词向量，CBOW和Skip-gram的示意图如图2.1所示。



图2.1 CBOW和Skip-gram示意图

CBOW模型是根据上下文的信息来预测中间的目标词。CBOW模型对NNLM模型做了简化：隐藏层不再是上文各词词向量的拼接，而是使用上下文各词词向量的平均值，减少了计算量；去掉了tanh隐藏层，提升了模型的训练速度。CBOW模型包含3层：输入层、投影层和输出层。输入层共有n-1个词的one-hot表示作为词向量，组成上下文的表示：公式1。投影层将输入层的n-1个向量做均值计算：公式2。输出层一共V个节点，第i个节点表示中心词是词wi的概率。CBOW模型根据上下文的表示，直接对目标词进行预测：公式3。对于整个语料库而言，CBOW的优化目标为最大化：公式4

Skip-gram模型与CBOW模型刚好相反，它是根据一个单词来预测它的上下文信息。Skip-gram模型训练一个双层神经网络，模型输入为一个词，来预测这个词的前n个词和后n个词出现的概率。Skip-gram和CBOW一样没有隐藏层。Skip-gram的优化目标为最大化：公式5

## 2.2.3 词向量的使用

目前可供使用的训练词向量的方法有很多，因为本文所使用的编程语言是Python，Python编程环境中有Gensim库可以直接调用models.word2vec来训练词向量。[Radford et al. 2017]提出使用任务相关的语料训练得到的词向量相对于别的词向量在特定任务中会有更好的效果。由于本文所涉及的项目是针对金融领域的文本进行信息抽取且命名实体识别的序列标注任务是针对每个字来进行标注的，所以使用金融相关文本作为训练字向量的文本语料，训练了一个包含50102个字的100维的字向量模型。

## 2.3 BERT模型

一般的词向量都是通过浅层网络进行无监督训练得到的，虽然在词的级别上有着不错的特性，但是缺少对连续文本的内在联系和语言结构的表达能力。因此希望通过大量数据来预训练一个大型的神经网络，用它来对文本进行特征提取之后再去做后续的任务，以期望得到更好的效果。

2018年Google提出的BERT模型结合了当时所有的预训练模型的优点，并采用了大量的数据和更大的模型来进行预训练，得到了当时在业界中表现最好的预训练模型。

BERT模型主要使用Transformer编码器作为主体模型结构。Transformer是一种完全舍弃了RNN的循环式网络结构的，完全基于注意力机制的网络结构模型。图2.2所示为Transformer编码器的结构图。

图2.2

Transformer编码器的核心思想是计算一句话中每个词对于这句话中所有词的相互关系，认为这些词与词之间的相互关系在一定程度上反映了这句话中不同词之间的关联性和重要程度，再利用这些相互关系来调整每个词的权重来获得每个词的新的表征。这个新的表征不但蕴含了词本身的含义，还蕴含了这个词与其他词的关系，相比于词向量，这样的词表征所包含的信息更加全面。Transformer就是通过对输入的文本不断进行这样的注意力机制层和普通非线性层交叠来得到最终的文本表征。

BERT模型主要的创新在于使用MaskLM的方式来训练语言模型以及增加句子级别的连续性预测任务“next sentence prediction”。MaskLM是指随机遮挡输入中的一些tokens并用一些特殊的符号进行替代，然后在预训练中对他们进行预测。BERT模型为了和后续任务保持一致，按一定比例在需要预测的词的位置上输入原词或者输入某个随机的词。由于一次输入的文本只有部分词被用来训练，所以BERT在效率上会较低，收敛会需要更多的训练步数；“next sentence prediction”是指预测输入文本是否为连续的文本。在训练时输入模型的第二个片段有50%的概率从全部文本中随机选择一句话，还有50%的概率选择第一段文本的后续文本。引入这个任务是为了让模型更好地学习连续的文本片段之间的关系。

## 2.4 双向长短时记忆模型

## 2.4.1 循环神经网络

循环神经网络（Recurrent Neural Networks，简称RNN）[Rumelhart et al. 1986]指的是一个序列当前的输出和之前的输出存在某种关联的神经网络。具体的表现形式为网络会对前面的信息进行记忆，保存在网络的内部状态中，并应用于当前输出的计算中。所以循环神经网络常常被用于对序列数据进行建模。循环神经网络的结构图如图2.3所示。

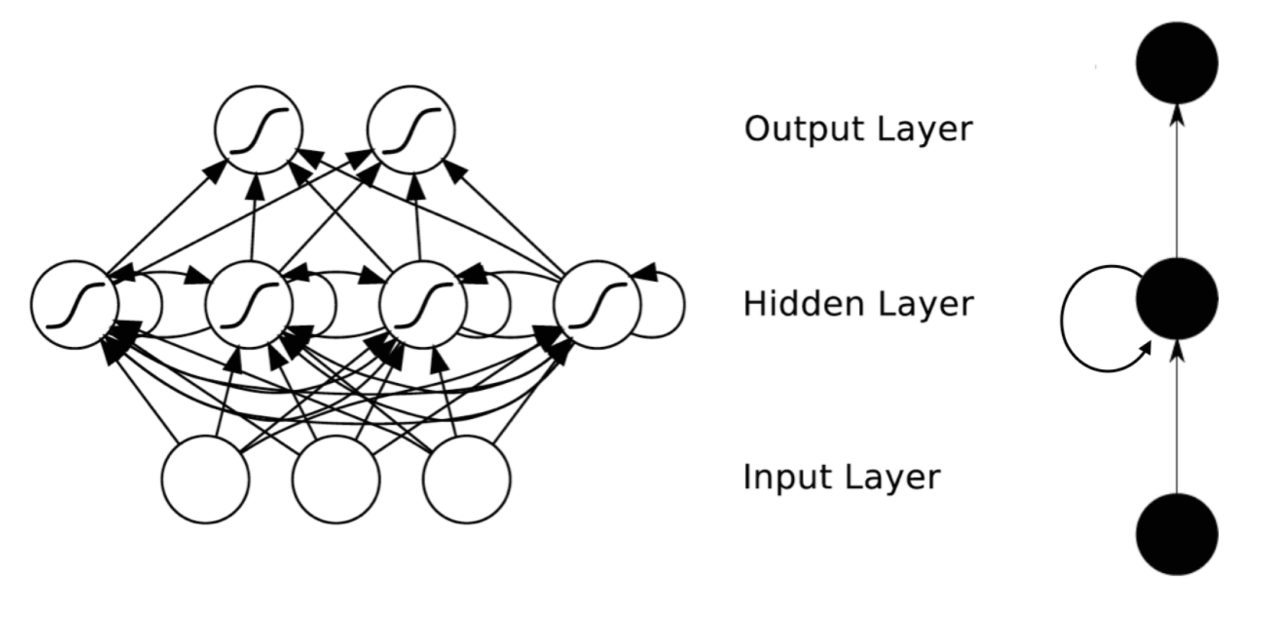


图2.3 循环神经网络的结构图

理论上循环神经网络能够对任何长度的序列数据进行处理，但是在实践中发现”梯度消失”会影响循环神经网络的训练，随着神经网络层数的增加，“梯度消失”的风险就会越大，当层数很大时，底层的神经元将接收不到返回的信号，出现“梯度消失”的问题。而长短时记忆模型（Long-Short Term Model，简称LSTM）[Hochreiter et al. 1997]的出现缓解了这个问题。

## 2.4.2 长短时记忆模型

长短时记忆模型的提出是为了解决循环神经网络的“梯度消失”问题，它之所以能够缓解这个问题主要取决于LSTM特有的门型结构，如图2.4所示。



图2.4 LSTM结构示意图

可以看到，LSTM主要包括遗忘门、输入门和输出门等结构，其中：

遗忘门：公式7

输入门：公式8

输出门：公式9

新记忆单元：公式10

最终记忆单元：公式11

隐藏层：公式12

三个不同的门型结构和记忆单元是LSTM的核心所在，记忆单元负责保存历史和当前的重要信息，遗忘门负责控制选择性丢弃之前的记忆单元中的一些信息，输入门负责控制当前的输入中的某些信息加入到记忆单元中，输出门负责控制记忆单元中的哪些信息可以进行输出。这些门控结构可以帮助LSTM在长时和短时的记忆中都能处理并解决数据依赖以及选择是否丢弃信息（保留重要信息并丢弃不重要的信息）。三个门相互合作是LSTM能够解决RNN梯度消失/梯度爆炸的关键所在。

## 2.4.3 双向长短时记忆模型

一般我们所说的RNN、LSTM都是单向的，也就是说在一个方向上进行隐藏状态的传递，然而在某些任务中需要考虑的不仅仅是当前输入单词前面的文本信息，还需要考虑当前输入单词后面的文本信息。双向长短时记忆模型就可以解决这种类型的问题，它不仅可以从前向后传递隐藏状态的信息，还可以从向前传递隐藏状态的信息，这样在计算输出时考虑的就是完整的上下文信息。可以把双向长短时记忆模型看成是两个方向相反的LSTM的叠加，在分别做计算后把两个状态向量拼接得到最后的输出。

## 2.5 深度残差网络

深度学习中对于网络深度加深遇到的主要问题是梯度消失和梯度爆炸，传统对应的解决方案是数据的初始化（normalized initialization）和（batch normalization）正则化。

深度残差网络（Deep Residual Network，简称DRN）[He K M. et al. 2015]的提出是源于一个“反常”的现象：在训练深度神经网络的时候，训练误差和测试误差都随着网络深度的增加而增加（Degradation）。在自然语言处理中，利用网络结构训练文本数据，浅层的网络更容易学习到文本中的词汇信息，而深层的网络更容易学到文本中的语义信息，但是由于层数的增加会导致训练误差的累积，残差网络将浅层的网络与深层的网络进行级联再输出，很好地控制了训练误差的累积，在自然语言处理的各项任务中取得了很好的效果。

假设深度网络中的某隐藏层为公式13，如果多个非线性层组合可以近似于一个复杂函数，那么同样可以认为隐藏层的残差近似于某个复杂函数，那么可以将隐藏层表示为公式13。这样一来就可以得到一种全新的残差结构单元，如图2.5所示。



图2.5 残差结构单元示意图

残差单元的输出由多个卷积层级联的输出和输入元素相加（维度相同的情况下），再经过ReLU激活后得到。将多个这样的结构级联起来，就得到了残差网络。

可以注意到残差网络有这样几个特点：1.网络较瘦，控制了参数数量；2.存在明显层级，特征图个数逐层递进，保证输出特征表达能力；3.使用了较少的池化层，大量使用下采样，提高传播效率；4.没有使用Dropout，利用BN和全局平均池化进行正则化，加快了训练速度。

## 2.6 远程监督

远程监督（Distant Supervision）是目前关系抽取中比较常见的一种做法。它既不是传统意义上的监督学习，也不是无监督学习，它属于一种用KB（Knowledge Base）去对齐朴素文本的标注方法：1.使用NET（Named Entity Tagger）标注；2.对已标注的金融领域文本中出现的三元组提取特征（从所有出现该三元组的句子中），构造训练数据；3.采用多类别逻辑回归进行分类。

在测试使用时，先使用命名实体识别模型对目标文本中的命名实体进行标注，抽取其中的命名实体对和特征。如果多个句子的命名实体对一样，则把它们的特征合并在同一个特征向量中，再利用逻辑回归分类器对关系名称进行识别。

简单来说，如果训练语料中的句子所包含的实体对在已有的KB中有关系的体现，则认为语料库中所包含相同实体对的句子都表达相应关系。虽然这样做可以减少关系抽取任务对人工标注数据的依赖，但是会引入很多噪声数据。因为两个实体之间可能有别的关系（不同于KB中已有的关系）或者是没有关系，这样的训练数据就会对我们的关系抽取任务产生一定影响。

近些年来，也有很多国内外学者在研究如何降低远程监督中的噪声数据对关系抽取造成的影响，如使用Multi-instance从训练集中抽取置信度较高的训练样例来训练模型、利用Attention机制对数据进行全方位的权重计算从而得到全面而不失“选择”的训练数据等方法。

本文在利用远程监督处理标注数据时，结合Attention机制和残差网络（ResNet）共同作用减少噪声数据对关系抽取模型带来的影响。

## 2.7 其他相关技术

本节主要介绍了除了模型外的一些项目中用到的技术，比如开源深度学习框架Tensorflow，Flask框架，Docker容器技术以及Kubernetes集群等。

## 2.7.1 深度学习框架Tensorflow

Tensorflow既是一个实现机器学习算法的接口，也是执行机器学习算法的框架。它前端支持Python、C++、Go、Java等多种开发语言，后端使用C++、CUDA等写成。除了执行深度学习算法，Tensorflow还可以用来实现很多传统机器学习算法，如逻辑回归、随机森林等。

Tensorflow使用数据流式图来规划计算流程，可以把计算映射到不同的硬件甚至是操作系统。Tensorflow的计算可以表示为有状态的数据流式图，可以让用户简单地实现并行计算，同时使用不用的硬件资源进行训练，同步或异步地更新全局共享的模型参数和状态。

## 2.7.2 Flask框架

Flask是一个由Python开发的轻量级Web框架。flask内核内置了两个重要的组件jinja2和Werkzeug，所有其它的组件都是通过易扩展的插件系统集成进来的。

Werkzeug本质是Socket服务端，用于接收http请求并对请求进行处理，之后触发Flask框架。Flask框架结合提供的功能对请求进行处理再返回给用户，如果要返回给用户复杂的内容时，要借助jinja2模板来实现对模板的处理。

Flask支持用扩展来给应用添加功能，包括数据库集成、表单验证、开放认证技术等。Flask旨在保持核心简单且易于扩展。

## 2.7.3 Docker容器技术

Docker是一个开源的引擎，可以轻松的为任何应用创建一个轻量级的、可移植的、自给自足的容器。开发者编译测试通过的容器可以批量地在生产环境中部署，包括VMs（虚拟机）、bare metal、OpenStack集群和其他基础应用平台。

Docker跟传统的虚拟化方式相比具有很多优势，其优势主要体现在：

1. 更高效的虚拟化：Docker容器的运行不需要额外的hypervisor支持，它是内核级的虚拟化，因此可以实现更高的性能和效率。
2. 更安全的运行应用：Docker容器内运行的进程完全与系统隔离，一些恶意行为对系统造成的影响不会波及宿主系统。
3. 更快速的交付和部署：开发者可以使用一个标准的镜像来构建一套开发容器，开发完成之后，运维人员可以直接使用这个容器来部署代码、Docker可以快速创建容器，快速迭代应用程序，并使整个过程全程可见，使团队其他成员更容易理解应用程序是如何创建和工作的。
4. 更轻松的迁移和扩展：Docker容器几乎可以在任意平台上运行，包括物理机、虚拟机、公有云、私有云、服务器等。这种兼容性可以帮助用户很轻松地将应用程序从一个平台直接迁移到另一个平台上。

## 2.7.4 Kubernetes集群

Kubernetes（简称k8s）是一个自动化容器操作的开源平台，这些操作包括部署、调度和节点集群间扩展。上一节中介绍的Docker容器就可以看作是Kubernetes内部使用的低级别组件。使用Kubernetes可以：

1. 自动化容器的部署和复制；
2. 随时扩展或收缩容器规模；
3. 将容器组织成组，并且提供容器间的负载均衡；
4. 易于应用程序容器的版本升级；
5. 提供容器弹性，可以替换失效的容器。

## 2.8 本章小结

本章主要介绍了本文对信息抽取中的命名实体识别和关系抽取的研究中涉及的相关技术。先是介绍了作为模型输入的深度学习自然语言问题的基础：词向量。之后又分别介绍了BERT模型、双向长短时记忆模型、深度残差网络、远程监督等，并介绍了本文实现模型使用的深度学习框架Tensorflow，实现信息抽取服务的flask框架、Docker容器技术、Kubernetes集群等。

# 

# 第三章 信息抽取服务分析与设计

## 3.1 项目总体规划

## 3.2 信息抽取服务需求分析

## 3.3 输入层

### 3.3.1 字向量

### 3.3.2 额外特征

## 3.4 网络层

## 3.5 输出层

### 3.5.1 全连接层

### 3.5.2 CRF层

## 3.6 实验

### 3.6.1 实验数据与采集

### 3.6.1.1 数据爬虫

### 3.6.1.1 实验数据

### 3.6.2 实验评价指标

### 3.6.3 模型训练的实现

### 3.6.4 模型参数设置

### 3.6.5 不同模型的性能对比

## 3.7 本章小结

# 

# 第四章 信息抽取服务的实现

## 4.1 模型的整体架构

## 4.2 输入层

### 4.2.1 字向量

### 4.2.2 相对位置向量

## 4.3 卷积层

## 4.4 残差卷积块

## 4.5 最大池化层

## 4.6 输出层

## 4.7 实验

## 4.7.1 实验数据

## 4.7.2 实验评价标准

## 4.7.3 模型训练的实现

## 4.7.4 模型参数设置

## 4.7.1 不同模型的性能对比

## 4.8 本章小结

# 第五章 实验对比与分析

宋体，小四

# 第六章 总结与展望

## 6.1 总结

本文以上海某大数据人工智能公司的知识图谱产品中的信息抽取模块为背景，介绍了该项目中信息抽取模块的两个子任务命名实体识别和关系抽取的研究现状，综述了基于深度学习的信息抽取任务所涉及的相关技术，包括：词向量、双向长短时记忆模型、深度残差网络、注意力机制和远程监督。阐述了文本所选择模型的详细设计和具体实现。

在模型选择方面，命名实体识别模型在对比了BiLSTM+CRF、CRF、LSTM、BiLSTM在数据集上的表现后，本文选择了性能较好的BiLSTM+CRF模型。此模型考虑了上下文信息的影响，也学习到了一些约束条件，有较好的分类效果；关系抽取模型是在对比了CNN、CNN+Attention和ResCNN在数据集上的表现后，选择了效果更好的ResCNN。ResCNN利用了残差学习的特性使网络结构学习到了更多的语言特性从而得到了更好的效果。最后，本文将两个模型固化成pb模型，可供后续进行预测使用。

## 6.2 进一步工作展望

由于项目的研发时间较短，所以关系抽取模型的效果相对来说还比较一般，希望在未来的时间里有机会在现有模型的基础上进行更多地尝试：

1. 加入Attention机制；
2. 利用PCNN代替CNN；
3. 利用新的损失函数。

# 

# 参 考 文 献

**[MUC-6, 1996] MUC-6, the Sixth in a Series of Message Understanding Conferences, was held in November 1996[OL].**<http://cs.nyu.edu/cs/faculty/grishman/muc6.html> **.**

**[Milleret al., 2000] Miller, Scott, Heidi Fox, Lance Ramshaw, and Ralph Weischedel. “A novel use of statistical parsing to extract information from text.” *In Proceedings of NAACL*, 2000.**

**[Bikel et al., 1999] Bikel D M,Schwarta R,Weischedel R M.An Algorithm that Learns What`s in a Name[J].*Machine Learning Journal Special Issue on Natural Language Learning*, 1999, 34(1-3): 211-231.**

**[Tsai et al., 2004] Tsai T,WU S,Lee C, et al. Mencius: A Chinese Named Entity Recognizer Using the Maximum Entropy based Hybrid Model[J]. *International Journal of Computational Linguistics & Chinese Language Processing*, 2004, 9(1):65-81.**

**[McCallum et al., 2003] McCallum A,Li W.Early Results for Named Entity Recognition with Conditional Random Fields, Features Induction and Web-enhanced Lexicons[C]. *In Proceedings of the 7th Conference on Natural Language Learning at HLT-NAACL*,2003: 188-191.**

**[张祝玉等, 2008] 张祝玉，任飞亮，朱靖波. 基于条件随机场的中文命名实体识别特征比较研究[C]. 见: *第4届全国信息检索与内容安全学术会议论文集*.2008.**

[Wu et al., 2015] Yonghui Wu, Min Jiang, Jianbo Lei,Hua Xu. Named Entity Recognition in Chinese Clinical Text Using Deep Neural Network. Stud Health Technol Inform. 2015;216:624-8.

[Z Huang et al., 2015] Zhiheng Huang, Wei Xu, Kai Yu. Bidirectional LSTM-CRF Models for Sequence Tagging. *arXiv*, 2015, 1508.01991 [cs.CL]

[Devlin J et al., 2018] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. *arXiv preprint arXiv*:1810.04805, 2018.

**[Kambhatla, 2004] Kambhatla, Nanda.”Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations.” *In Proceedings of ACL*, 2004.**

**[Zhao and Grishman, 2005] Zhao, Shubin, and RalphGrishman. Extracting relations with integrated information using kernel methods. *In Proceedings of ACL*, 2005.**

[Culotta et al., 2006] Culotta, Aron, Andrew McCallum, and Jonathan Betz. Integrating probabilistic extraction models and datamining to discover relations and patterns in text. *In Proceedings of HLT-NAACL*, 2006.

[Mintz et al., 2009] Mintz, Mike, Steven Bill, RionSnow, and Dan Jurafsky. Distant supervision for relation extraction without labeled data. *In Proceedings of ACL-IJCNLP*, 2009.

[Socher et al. 2012] Socher , Richard, et al. Semantic compositionality through recursive matrix-vectorspaces. *In Proceedings of EMNLP-CoNLL*, 2012.

[Zeng et al., 2014] Daojian Zeng, Kang Liu, et al. Relation classification via Convolutional Deep Neural Network. *In Proceedings of COLING*, 2014

[Santos et al., 2015] Cicero Nogueira dos Santos, Bing Xiang, Bowen Zhou. Classifying Relations by Ranking with Convolutional Neural Networks. *In Proceedings of ACL*, 2015.

[Miwa er al., 2016] Makoto Miwa, Mohit Bansal. End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures. *In Proceedings of ACL*, 2016.

[Lin et al., 2016] Yankai Lin, Shiqi Shen, Zhiyuan Liu, et al. Neural Relation Extraction with Selective Attention over Instances. *In Proceedings of ACL*, 2016.

[Bengio et al., 2002] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. *IEEE Trans Neural Netw*, 2002, 5(2):157-166.

[Hinton, 1986] Hinton G E. Learning distributed representations of concepts[C]. *Eighth Conference of the Cognitive Science Society*. 1986.

[Harris, 1954] Harris Z S.Distributional structure[J]. *Word*, 1954, 10(2-3): 146-162.

[Pennington et al., 2014] Jeffrey Pennington, Richard Socher, and Christopher D. GloVe: Global Vectors for Word Representation. *Manning*. 2014.

[Bengio et al., 2006] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model[J]. *Journal of Machine Learning Research*, 2006, 3(6):1137-1155.

[Mikolov et al., 2013a] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[J]. *Computer Science*, 2013.

[Mikolov et al., 2013b] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]. *International Conference on Neural Information Processing Systems*. Curran Associates Inc. 2013:3111-3119.

[Radford et al., 2017] Radford A, Jozefowicz R, Sutskever I. Learning to Generate Reviews and Discovering Sentiment[J]. 2017.

[Rumelhart et al., 1986] David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams. Learning representations by back-propagating errors[J]. *Nature*, 1986, 323(3):533-536.

[Hochreiter et al., 1997] Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural computation*, 1997, 9(8):1735-1780.

[He K M et al., 2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren, et al. Deep Residual Learning for Image Recognition[J]. *arXiv*, 2015, 1512.03385.

**[Hinton et al., 2012]** Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. *arXiv preprint arXiv*:1207.0580, 2012.

# 致 谢

转眼之间我的研究生生涯就要结束了，对南大有一丝丝不舍，对未来有一丝丝憧憬和期待，在此我要向南大的研究生老师们和同学们表达最真挚的感谢。

首先我要感谢我的毕业论文指导老师伏晓老师，感谢伏老师对我毕业论文的指导和修正。毕业论文开题前，伏老师与我就毕业论文事宜进行了详谈，给我毕业论文的选题和方向做出了指导和建议，在论文的写作过程中，伏老师一直耐心地给我写作意见，伏老师对毕业论文负责任的态度，给了我莫大的帮助。

其次我要感谢南京大学软件学院的所有老师们，感谢老师们严谨的教学态度和学术态度，帮助我培养了良好的学习习惯和学习态度。还要感谢周围同学们在我学习生活中给予的帮助，给了我继续前进的动力。

再次我要感谢我的家人，他们在我最困难的时候默默地支持和鼓励着我，让我战胜了困难，最终进入南京大学就读研究生，实现了自己的梦想。

我还要感谢所有在我实习的时候热心帮助我的同事们，在实验的时候有很多不懂的地方都是靠着同事们的帮助帮助我踩过一个又一个坑。

最后，由衷地感谢在百忙之中评阅论文的各位专家和教授。

# 版权及论文原创性说明

任何收存和保管本论文的单位和个人，未经作者本人授权，不得将本论文转借他人并复印、抄录、拍照或以任何方式传播，否则，引起有碍作者著作权益的问题，将可能承担法律责任。

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含其他个人或集体已经发表或撰写的作品成果。本文所引用的重要文献，均已在文中以明确方式标明。本声明的法律结果由本人承担。

作者签名：

日期： 年 月