

文章编号: 1003-0077(2019)12-0001-18

基于深度学习的关系抽取研究综述

庄传志^{1,2}, 靳小龙^{1,2}, 朱伟建^{1,2}, 刘静伟^{1,2}, 白 龙^{1,2}, 程学旗^{1,2}

(1. 中国科学院 计算技术研究所 中国科学院网络数据科学与技术重点实验室, 北京 100190;

2. 中国科学院大学 计算机与控制学院, 北京 100049)

摘 要: 关系抽取(RE)是为了抽取文本中包含的关系,是信息抽取(IE)的重要组成部分。近年来,研究人员利用深度学习技术在该领域开展了深入研究。由于神经网络类型丰富,基于深度学习的关系抽取方法也更加多样。该文从关系抽取的基本概念出发,对关系抽取方法依据不同的视角进行了类别划分。随后,介绍了基于深度学习的关系抽取方法常用的数据集,并总结出基于深度学习的关系抽取框架。在此框架下,对关系抽取方法在面向深度学习的输入数据预处理、面向深度学习的神经网络模型设计等方面的具体工作进行了分析与评述,最后对未来的研究方向进行了探讨和展望。

关键词: 关系抽取;深度学习;远程监督;联合学习

中图分类号: TP391

文献标识码: A

Deep Learning Based Relation Extraction: A Survey

ZHUANG Chuanzhi^{1,2}, JIN Xiaolong^{1,2}, ZHU Weijian^{1,2}, LIU Jingwei^{1,2}, BAI Long^{1,2}, CHENG Xueqi^{1,2}

(1. CAS Key Laboratory of Network Data Science and Technology,

Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;

2. School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Information Extraction(IE) is a task of natural language processing that involves extracting structured information from plain unstructured text. Relation Extraction(RE) is a crucial component in IE. Recently, researchers pay great attention to the method of deep learning, resulting various methods in this field. Starting from the basic concept of relationship extraction, this paper groups methods from different perspectives, introduces the popular data sets, and outlines the deep learning framework for relationship extraction. This paper analyzes and reviews the details in data preprocessing and model design in these methods. Finally, the future research direction is discussed.

Keywords: relation extraction; deep learning; distant supervision; joint learning

0 引言

伴随着互联网、云计算等技术的发展,产生了海量的文本数据信息,其中包含大量有价值的知识。然而由于这些信息内容的多元异质、组织松散,给知识利用带来很多困难。信息抽取的目的就是将非结构化或半结构化描述的自然语言文本转化成结构化内容以更好地利用知识,包括抽取指定类型的实体、关系、事件等信息。关系抽取是信息抽取的重要子任务,主要负责从无结构文本中识别出实体,并抽取

实体之间的语义关系,被广泛用在信息检索^[1]、问答系统^[2]中。

深度学习是机器学习中的一个新领域,由 Hinton 等人于 2006 年提出,其概念源于人工神经网络的研究。深度学习大量借鉴关于人脑、统计学和应用数学的知识,通过组合低层特征形成更加抽象的高层特征表示,以发现数据的分布式特征表示^[3]。近年来,得益于更强大的计算机、更大的数据集和能够训练更深的技术,深度学习的普及性和实用性都有了极大的发展,并成为机器学习领域的研究热点,在图像处理、语音处理等应用领域均取得突破性的

收稿日期: 2019-04-22 定稿日期: 2019-08-30

基金项目: 国家重点研发计划(2016YFB1000902);国家自然科学基金(61772501,61572473,61572469,91646120)

进展,并已经成为自然语言处理领域的重要研究工具,被广泛用在词性标注^[4]、命名实体识别^[5]、语义角色标注^[6]等领域。

由于社会各界对知识图谱的深入研究,关系抽取作为知识图谱构建的重要环节引起了广泛关注。很多学者将深度学习用于关系抽取领域的研究中,取得了突出成绩^[7-11]。近年来,基于神经网络的方法较多,一些学者做了梳理,如 Kumar 等^[12]介绍了卷积神经网络在该领域的部分研究进展,但缺少对整个领域开展的系统深入的梳理与总结工作。

为此,本文对基于深度学习的关系抽取方法进行了归纳总结。首先引入关系抽取简介,详细介绍方法的分类以及基于深度学习的关系抽取方法常用数据集,并提出基于深度学习的抽取框架。据此框架,本文详细阐述了深度学习在关系抽取领域的研究进展,并展望未来发展方向。本文集中于英文语料的二元关系抽取的研究讨论,多元关系抽取和其他语言的关系抽取不在本文讨论的范围。

1 关系抽取简介

关系抽取的定义和分类较多,一直以来缺少详细的介绍,容易造成概念上的混淆。在介绍关系抽取方法之前,本节首先介绍关系抽取的基本概念、方法分类和基于深度学习的关系抽取方法常用的数据集,并给出基于深度学习的关系抽取框架。

1.1 关系抽取基本概念

完整的关系抽取包括实体抽取和关系分类两个子过程。实体抽取子过程也就是命名实体识别,对句子中的实体进行检测和分类;关系分类子过程对给定句子中两个实体之间的语义关系进行判断,属于多类别分类问题。

例如,对于句子“青岛坐落于山东省的东部”,实体抽取子过程检测出这句话具有“青岛”和“山东”两个实体。关系分类子过程检测出这句话中“青岛”和“山东”两个实体具有“坐落于”关系而不是“出生于”关系。在关系抽取过程中,多数方法默认实体信息是给定的,那么关系抽取就可以看作是分类问题。

1.2 关系抽取方法的分类

目前,常用的关系抽取方法有 5 类,分别是基于模式匹配、基于词典驱动、基于机器学习、基于本体和混合的方法^[13]。基于模式匹配和词典驱动的方法

法依靠人工制定规则,耗时耗力,而且可移植性较差,基于本体的方法构造比较复杂,理论尚不成熟^[14]。基于机器学习的方法以自然语言处理技术为基础,结合统计语言模型进行关系抽取,方法相对简单,并具有不错的性能,成为当下关系抽取的主流方法,下文提到的关系抽取方法均为机器学习的方法。

可从多个视角对基于机器学习的关系抽取方法进行划分,比如根据文本抽取的范围将其分为句子级关系抽取和篇章级关系抽取两类,根据抽取目标将其分为句子级关系抽取和实体对级关系抽取两类。

为了清晰地介绍基于深度学习的抽取方法,本文从训练数据的标记程度、使用的机器学习方法、是否同时进行实体抽取和关系分类子过程以及是否限定关系抽取领域和关系专制四个角度对机器学习的关系抽取方法进行分类。

1.2.1 根据训练数据的标记程度分类

根据训练数据的标记程度可以将关系抽取方法分为有监督、半监督和无监督三类^[15],近年来,远程监督被引入到关系抽取任务中,并成为关系抽取的一类重要方法。

有监督学习处理的基本单位是包含特定实体对的句子,每一个句子都有类别标注。此类方法将关系抽取当作多类别分类问题来解决,首先根据训练数据来获取句子的有效表示,然后采用特定的方法来比较不同句子之间的语义相似性,并利用该模型对测试数据进行预测。有监督的关系抽取能够有效利用样本的标记信息,准确率和召回率都比较高,是关系抽取领域主流的研究方法。

然而,有监督的关系抽取方法的最大缺点是需要大量的人工标记训练语料,代价较高,因此远程监督被用于关系抽取。远程监督基于这样一种假设:如果两个实体在知识库中存在某种关系,那么特定语料中所有包含这两个实体的非结构化句子均能表示这种关系。远程监督利用已有知识库中的知识对语料进行标注,有效解决了关系抽取中的数据标注问题,Mintz 等^[16]最早将 Freebase 中的三元组对齐到《纽约时报》新闻文本语料库中,便捷地得到大量有标注训练数据,此类数据集被称为远程监督数据集。

远程监督关系抽取方法处理的基本单位是由包含特定实体对的多个句子所构成的句包,对数据的标注是句包级别的标注,并未对句包中的每一个句

子进行明确标注,因此可以将其看作是针对句包的有监督学习。

半监督学习中,句子作为训练数据的基本单位,只有部分是有类别标注的。半监督的学习方法主要采用自助、协同训练等方法进行关系抽取^[17]。对于要抽取的关系,该方法首先通过人工为每种关系设定若干种子实例,然后迭代地在语料库中获得更多相关句子集合并抽取表达关系的模板和实例。此类方法让学习器不依赖外界交互,自动地利用未标记样本来提升学习性能。

无监督学习中,完全不需要对训练数据进行标注,此类方法包含实体对标记、关系聚类 and 关系词选择三个过程。无监督学习假设相同语义关系的实体对具有相似的上下文信息,因此利用每个实体对对应的上下文信息来代表实体对的语义关系,并对所有实体对的语义关系进行聚类^[18]和选择。

本文介绍的基于深度学习的关系抽取方法均属于有监督学习或远程监督学习方法。

1.2.2 根据使用的机器学习方法分类

根据使用机器学习方法不同,可以将关系抽取划分为三类:基于特征向量的方法、基于核函数的方法以及基于神经网络的方法。基于特征向量的方法和基于核函数的方法,一般称为传统的关系抽取方法。

基于特征向量的方法通过从包含特定实体对的句子中提取出语义特征,构造特征向量,然后通过使用支持向量机^[19]、最大熵^[20]、条件随机场^[21]等模型进行关系抽取。基于核函数的方法充分利用句子的特定组织形式^[22],可以是字符串序列或树结构,不需要显式定义复杂的语义特征,通过设计特定核函数来计算句子之间的相似度,根据相似度进行分类。

对于基于特征向量的方法而言,这些特征多需要人工设计,最终选择哪些特征往往通过启发式的试验得出,因工作量较大,并且其所使用的实体类型、词性标签、依存关系等特征,已经覆盖了大多数可利用的语言现象,在分类效果方面很难进一步提升。对于基于核函数的方法而言,其重点是巧妙地设计核函数来计算不同关系实例特定表示之间的相似度,而如何设计核函数需要大量的人类工作,不适用于大规模语料上的关系抽取任务。

针对上述问题,很多学者开始将深度学习应用到有监督和远程监督关系抽取中,通过构造不同的神经网络模型来自动学习句子的特征,减少了复杂

的特征工程以及领域专家知识,具有很强的泛化能力。

1.2.3 根据是否同时进行实体抽取和关系分类子过程进行分类

根据是否在同一个模型里开展实体抽取和关系分类,可以将关系抽取方法分为流水线(pipeline)学习和联合(joint)学习两种。流水线学习是指先对输入的句子进行实体抽取,将识别出的实体分别组合,然后再进行关系分类,这两个子过程是前后串联的,完全分离。联合学习是指在一个模型中实现实体抽取和关系分类子过程。

流水线学习让每一个子过程都更灵活,使关系抽取更容易,因此多数基于深度学习的关系抽取方法都默认实体对是给定的,工作重点集中在关系分类子过程。然而此类方法忽视了两个子任务之间存在的关系,比如“小明跳槽去了华为”,“小明”是人名实体,“华为”是公司实体,有助于确定两者具有“受雇于”关系,而不是“出生于”关系;同样,这句话具有“受雇于”关系可以帮助确定“小明”是人名实体,“华为”属于公司实体。

联合学习的方法通过使两个子过程共享网络底层参数^[23]以及设计特定的标记策略^[11]来解决上述问题,其中使用特定的标记策略可以看作是一种序列标注问题。

1.2.4 根据是否限定关系抽取领域和关系类别分类

根据是否限定抽取领域和关系类别,关系抽取方法可以划分为预定义抽取和开放域抽取两类。预定义关系抽取是指在一个或者多个固定领域内对实体间关系进行抽取,语料结构单一,这些领域内的目标关系类型也是预先定义的。开放域关系抽取不限定领域的范围和关系的类别。现阶段,基于深度学习的关系抽取研究集中于预定义关系抽取。

1.3 常用数据集

基于深度学习的关系抽取方法常用的数据集有ACE关系抽取任务数据集、SemEval2010 Task 8数据集、NYT2010数据集等。

1.3.1 ACE关系抽取任务数据集

ACE(automatic content extraction)评测会议^[24]由美国国家标准技术研究院组织,其中一项重要评测任务就是实体关系识别。评测会议的各类任务语料来源包括网络上的专线新闻和广播新闻、新闻组、报纸新闻和电视对话、网络日志等。其中ACE2005关系抽取数据集包含599篇与新闻和邮

件相关的文档,其数据集内包含 7 大类 25 小类关系。

1.3.2 SemEval2010 Task 8 数据集^[25]

该数据集包含 9 种关系类型,分别是 Component-Whole、Instrument-Agency、Member-Collection、Cause-Effect、Entity-Destination、Content-Container、Message-Topic、Product-Producer 和 Entity-Origin。考虑到实体之间关系的方向以及不属于前面 9 种关系的“Other”关系,共生成 19 类实体关系。其中训练数据 8 000 个,测试数据 2 717 个。

1.3.3 NYT2010 数据集^[26]

NYT2010 数据集是 Riedel 等人在 2010 年将 Freebase 知识库中的知识“三元组”对齐到“纽约时报”新闻中得到的训练数据。该数据集中,数据的单位是句包,一个句包由包含该实体对的若干句子构成。其中,训练数据集从《纽约时报》2005—2006 年语料库中获取,测试集从 2007 年语料库中获取。

1.4 基于深度学习的关系抽取框架

基于深度学习的关系抽取框架由输入预处理、数据表示、网络模型学习 3 个部分组成,如图 1 所示。每一部分的输出作为下一部分的输入。在输入预处理部分,将整个句子或者该句子的特定范围作为神经网络的输入,并使用自然语言处理工具获取数据的显式特征;在数据表示部分,获取上述输入对应的低维向量表示;在网络模型学习部分,针对之前的输入范围和特征信息,设计网络模型,获取句子或者句包的表示,并根据训练数据进行模型的学习。

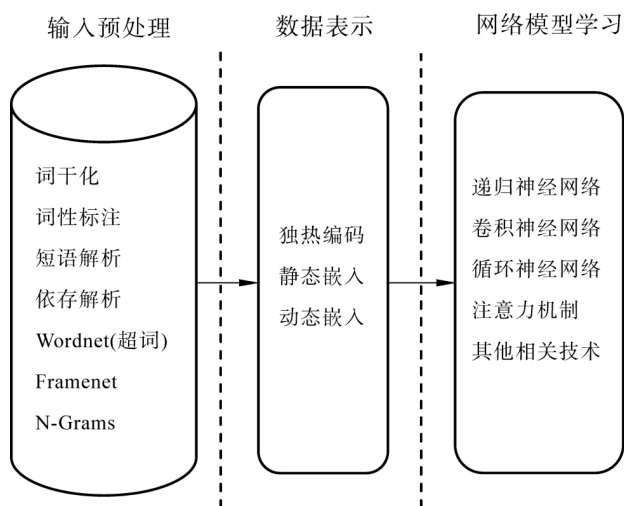


图 1 基于深度学习的关系抽取方法框架

基于深度学习的关系抽取,主要思想是选择样本中特定输入信息并设计合理的网络结构,将所要

进行抽取的样本转换成独特的向量表示来进行关系抽取。其中,有监督的关系抽取学习的是句子的表示,远程监督的关系抽取学习的是句包的表示,因此,远程监督的关系抽取还要设计特定的方法将句子的表示转换成句包的表示。这部分工作也是在网络模型构建过程中完成。

近年来研究人员主要从输入预处理和网络模型设计两部分入手,开展了广泛深入的研究。本文在第 2 节,详细介绍面向深度学习的输入数据预处理研究工作;第 3 节详细介绍面向深度学习的关系抽取方法在网络模型构建方面的工作;第 4 节介绍了基于深度学习关系抽取方法未来研究重点,并进行展望。最后对本文进行总结。

2 面向深度学习的输入数据预处理

深度学习模型主要根据语言学知识,构建特定的输入,根据形式的不同可将输入分为序列结构和树结构两大类。在数据预处理阶段,研究人员的工作集中在根据自然语言处理的相关知识,选择特定的输入范围,并根据句子的特征将其处理成特定的结构。

2.1 输入语句的范围

一般认为,将整个句子作为输入,包含的信息量最大,不至于丢失任何信息^[7,27]。然而,使用整个句子作为输入时,却可能包含冗余特征。尽管神经网络具有自动学习潜在特征的能力,研究人员仍然尽可能地在将语句送入神经网络之前去除不相关信息。在该阶段,研究人员采用了不同范围的输入来表示句子中两个实体所具有的关系信息。

Santos 等^[28]研究发现大部分关系可以由句子中两个实体之间的关键词反映出来。例如,如果实体之间具有“实体—目的地(实体 1, 实体 2)”关系,那么两个实体之间很有可能有方向词“到”或者“去”;如果实体之间具有“原因—结果(实体 1, 实体 2)”关系,那么实体 1 和实体 2 之间很有可能有如“导致”或“由……引起”等词语,并且这些词语距离实体较近。Rink 等^[29]认为对两个实体关系进行抽取,最重要的特征来自于两个实体、两个实体之间的单词以及两个实体周围的单词信息。

根据上述发现,研究认为选择局部关键信息作为输入可有效降低冗余信息。Hou 等^[30]仅仅考虑实体周边特定范围的词汇,如把两个实体前后 n 个

词,作为输入;Santos 等^[28]仅使用两个实体及它们之间的子句来抽取句子的信息;Zheng 等^[31]显式引入实体前后的词来抽取实体的信息,以此作为补充来提升抽取效果。

为了验证卷积神经网络(convolution neural network, CNN)模型在短文本上的效果, Qin 等^[32]使用比两个实体之间的子句多前后两个单词的子句作为输入。考虑到句子太长对关系抽取结果的影响, Liu 等^[33]选择输入的句子长度不超过一个特定的范围。

另一方面,利用句法分析的结果作为输入也可以有效地丰富句子的表示,揭示实体对之间的语义关系。句法分析分为成分句法分析和依存句法关系分析两种。

成分句法分析,以获取整个句子的句法结构或者完全短语结构为目的,把句子组织成短语组合的形式。Hashimoto 等^[34]使用整个句子的成分解析树作为输入;而 Socher 等^[8]认为,把句法成分树上两个实体之间的共同节点所在的子树作为输入,足够表达出实体对之间的语法关系。

依存句法分析主要诠释词间的依赖关系,可以识别句子中的“主谓宾”、“定状补”等语法成分,聚焦于句子的行为和主题。为了全面利用依存树上的语义信息, Yang 等^[35]使用整个句法依存树作为输入,并重新定义当前节点的上下文信息为其依存树上的父节点和子节点。

研究认为^[9],两个实体之间的最短依存路径可以描述两个实体之间的关系。因此, Ebrahimi 等^[36]使用两个实体之间的最短路径上所有节点所在的子树作为输入,既保留了关系的关键词,又减少了冗余信息的输入。为了最大限度利用最短依存树上的语义信息, Xu 等^[37]还使用了两个实体之间的最短依存路径上的箭头方向、路径上的关系标记等内容。

Nguyen 等^[38]指出,仅仅使用两个实体之间的最短路径作为输入,会忽略掉最短路径上关键词的子树信息。针对这个问题, Liu 等^[39]根据依存树给最短依存路径上的每个节点附加其依存句法树上的子树作为神经网络模型的输入,称之为依存增强子树。

同时, Santos 等^[28]和 Qin 等^[32]将额外的标记信息作为输入的组成部分。例如,由于训练样本中已使用实体标签标记目标实体,则在句子训练的过程中将[e1]\ [\e1]\ [e2]\ [\e2]看作是四个单词,有效提升了抽取效果。

2.2 输入语句的特征

研究人员引入神经网络的初衷是为了尽量避免对句子进行词性标注、句法分析等操作,减少人工构造特征以及自然语言处理工具的错误带来的误差累积和传播。然而后续研究发现,在语义信息不完备的情况下,人工构造的显式特征在神经网络模型中对于关系抽取问题仍具有一定的促进作用,因此,研究人员也使用多种特征的混合来表示单词,主要可以分为两大类,词汇特征和相对位置特征。

2.2.1 词汇特征

词汇特征指通过对句子使用一系列的自然语言处理工具和外部知识库提取的特征,是与语义相关的特征。常用的词汇特征有词性、命名实体类别、语法关系、Wordnet 超词等。

由于同一个单词在不同句子中的词性不同,则其含义也有所变化,故 Hashimoto 等^[34]使用了词性特征来表示单词。命名实体类别也可有助于理解实体之间的关系,例如,人名一地名更可能是实体一目的地关系,而不是通信—主题关系,故 Liu 等^[33]使用命名实体特征来表示单词。Wordnet 是一部语义词典,该词典有利于在具有相似概念的不同单词间构筑链接。例如,“parrot”这个词的上位词就是“bird”,这样不同的单词在 Wordnet 中有可能有相同的上位词,可以引入更多语义信息,因此 Xu 等^[10,37]使用 Wordnet 上位词来表示单词特征。

2.2.2 相对位置特征

相对位置特征是指当前单词与实体之间的距离。一般认为,在句子中,单词与实体的关系越近,其与对应关系的关联越重要。根据输入范围和内容的不同,可以将位置特征分为基于句子的位置特征和基于依存树的位置特征两大类。

基于句子的位置特征最早由 Collbert 等^[40]用来进行语义角色标记,随后 Zeng 等^[7]将位置特征引入到关系抽取中。句子中包含两个实体,因此每一个单词的位置特征至少有两个,例如,在句子“Convulsions that occur after DTap are caused by a fever”中,caused 距离 Convulsions 和 fever 的距离分别是 6 和 -3,这样一组数字就代表这个单词的位置信息。

位置特征在关系抽取中有一定的作用,然而也面临一些问题,例如,两个实体间的副词和不相关实体改变了实体间的相对距离,却不提供任何有用的信息,导致抽取效果降低。依存树能够缩短标记实

体之间的距离,并且可以帮助去掉冗余单词,因此很直接地想到把基于句子的位置特征迁移到依存树上。

Yang 等^[35]在依存树上设计了两类位置信息,分别记为 tpf1 和 tpf2。图 2 详细标识了当前依存树上每个节点距离其中一个实体[Convulsions] 的距离特征。

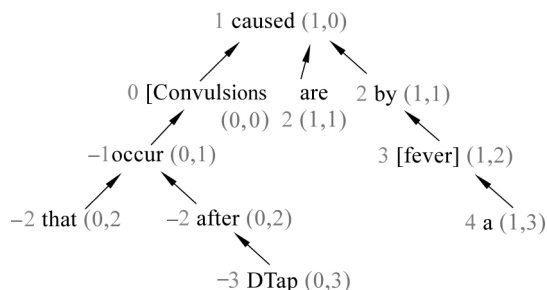


图 2 基于依存树的位置特征示意图^[35]

tpf1 用来表示依存树上的单词距离标记实体的相对距离,如果该单词为其中一个标记实体的子节点,那么该单词的 tpf1 值为负数,否则为正数,如图 2 中节点左侧的数字所示。tpf2 使用元组表示一个单词和对应标记实体分别到其最小共同节点的距离,如图 2 中节点的右侧标记。

为了丰富位置特征,Zhang 等^[41]根据依存树上两个实体之间的最短路径,还设计了以下三类相对依存位置特征:相对于根节点的特征(是否为根节点,是否为根节点的子节点,或者其他节点),相对于实体 e1 的特征(是否为 e1 节点,是否为 e1 的子节点,其他节点),相对于 e2 的特征(是否为 e2 节点,是否为 e2 的子节点,其他节点)。

通过上述分析可以看到,不同的输入范围的含义是不同的,给关系抽取造成的影响也不相同,通过预处理抽取合适的句子特征有利于提升抽取效果。

3 面向关系抽取的深度学习模型构建

基于深度学习的关系抽取方法模型构建的重点在于利用不同神经网络的特点来抽取样本的特征,以学习样本的向量表示。在学习过程中,根据所用的神经网络基本结构的不同,可将基于深度学习的关系抽取方法分为基于递归神经网络(recursive neural network, Rec-NN)的方法、基于卷积神经网络的方法、基于循环神经网络(recurrent neural network, RNN)的方法和基于混合网络模型的方法四类。

3.1 基于递归神经网络的关系抽取

在自然语言处理领域,句子的组合性是指短语是由一系列相邻词汇组合而成的,复杂句子的含义由其子句或短语根据特定的规则递归组合确定^[42]。递归神经网络是一种全连接的前向神经网络,其用于自然语言处理的重要原因就是可以依据特定的规则对句子成分进行组合。

基于递归神经网络的关系抽取方法首先利用自然语言处理工具对句子进行处理,构建特定的二叉树,然后解析树上所有的相邻子节点,以特定的语义顺序将其组合成一个父节点,如图 3 所示。这个过程递归进行,最终计算出整个句子的向量表示。向量计算过程可以看作是将句子进行一个特征抽取过程,该方法对所有的邻接点采用相同的操作。

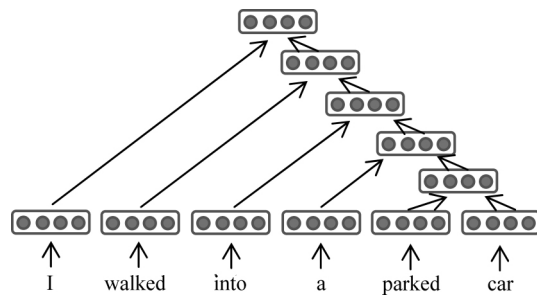


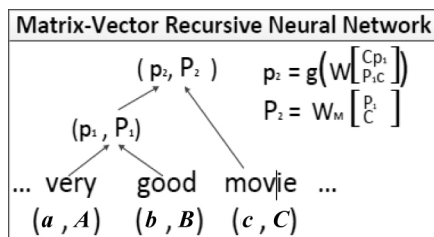
图 3 递归神经网络示意图^[43]

Socher 等^[44]最早提出这种递归神经网络,并将其用到情感分类中,取得了较好的效果。然而单词在组合成短语的过程中,不同的单词具有的意义不同,而且在改变其他单词或短语意义上的作用也千差万别。

针对自然语言的这个特性,Socher 等^[8]提出了 MV-RNN 方法。MV-RNN 为句法成分解析树上的每个节点分别定义了一个元组(向量,矩阵),向量代表节点本身的含义,矩阵代表节点具有的改变与之结合的节点的词义的能力。

具体如图 4 所示,这个方法的输入是句法成分解析树上两个实体到它们的最小共同节点构成的子树,其中 a 和 b 表示该节点的向量, A 和 B 表示该节点的矩阵。每一个非终端节点都要通过组合其子节点的语义得到对应的向量和矩阵。该子树的最终向量,通过自底向上地使用递归神经网络组合得到,作为句子的关系向量。

针对同一短语词性也可能具有不同词性这个特点,Hashimoto 等^[34]使用句法成分解析树上包含两个实体的最小子树作为输入,句法树上的叶子节点

图4 MV-RNN 网络结构示意图^[8]

的输入不再是单词,而是单词—词性对。在递归组合过程中,对每个子节点设计了不同的组合权重,权重的大小由当前节点的词性、短语类别、是否是句法头、是否位于两个实体之间句法树路径上等因素决定;在组合时,使用的组合矩阵也根据子节点—词性对来确定。

与句子的成分解析树相比,依存解析树可以对两个实体之间的更丰富的关系进行编码。针对这个特性,Ebrahimi 等^[36]使用依存解析树上两个实体之间的最短依存路径来构造二叉树,并适用于递归神经网络模型。

Ebrahimi 等^[36]采用两种方法构造二叉树。一种是基于启发式的固定的树结构,这种方法从叶子节点出发,递归地组合从属词—支配词对,构造一个

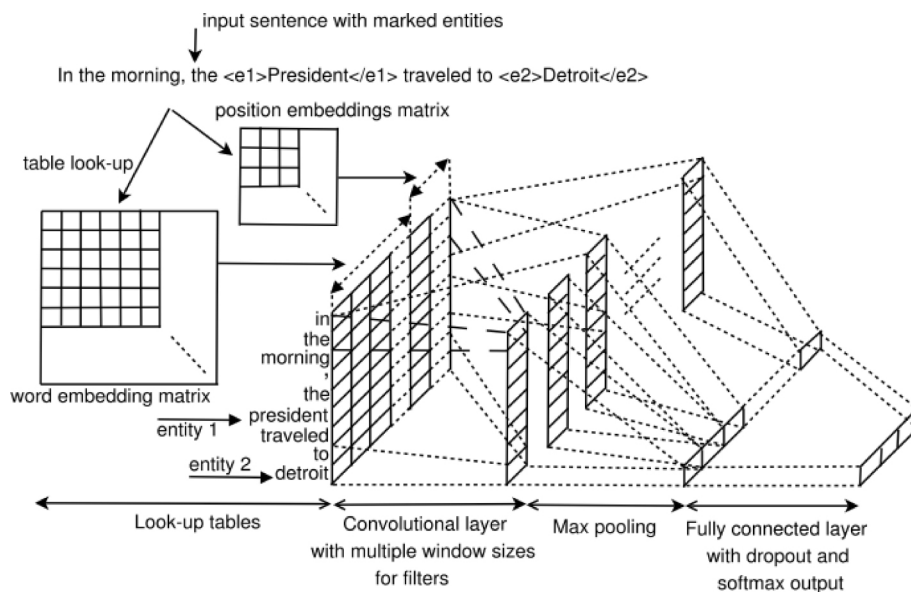
满二叉树,得到一个共同的祖先,这个祖先代表该句子的关系向量。第二种通过递归自编码的方法进行树的构建,目标是最小化树上原始输入节点之间的重构误差,找到最可能的树结构,最终根据构建的树来表示最终的节点。

基于递归神经网络的关系抽取方法的关键在于如何构造解析树,以及组合函数的选择。这里的神经网络可以有多层,触发函数也可以有多种选择。

3.2 基于卷积神经网络的关系抽取

卷积神经网络是当今计算机视觉系统的核心技术,在图像分类领域做出了巨大贡献。近些年,卷积神经网络被应用于各类自然语言处理任务,如语义解析、搜索检索、句子分类和语义角色标注等方面,均取得了较好成果。

基于卷积神经网络的关系抽取方法接受一个特定的向量矩阵作为输入,通过卷积层和池化层的操作将输入转换成一个固定长度的向量,并使用其他特征进行语义信息汇总,再进行抽取。基于卷积神经网络的关系抽取方法框架如图5所示,除了输入层、数据表示层之外,还有窗口层、卷积层、池化层、语义信息汇总层、分类层。

图5 基于卷积神经网络的关系抽取框架图^[36]

窗口层利用了语义分布假设的思想^[42],即:使用当前单词本身及单词的上下文,可以更清晰地表达当前单词的含义。卷积层通过多个卷积核对句子进行从上到下的卷积操作,来收集句子的局部的抽象特征。在实际操作中,窗口层一般通过设置卷积

核的大小来完成。池化层的作用在于对卷积层特征做进一步筛选,用来抽取句子级别的全局特征。语义信息汇总层是将池化层获取的特征与部分显式构造的特征结合,作为句子的最终向量表示。最后将这个向量送入到一个全连接网络进行分类。

Liu 等^[33]首次使用卷积神经网络来做关系抽取任务,对于多特征组成的特征向量输入,首先使用卷积层进行卷积操作,将输出进行池化操作,使得最终得到的句子向量表示与句子的长度无关。最后将该结果作为句子的向量表示送入到分类层中。

Zeng 等^[7]引入位置特征,并使用预训练的词向量进行关系抽取。在语义信息汇总层,Zeng 等人将获取的句子两类语义特征组合作为句子的最终向量表示,一类是通过对句子使用窗口层和卷积层操作获取句子语义特征,另一类是显式地设计词汇特征,包括两个实体本身、两个实体的前后单词以及两个实体在 Wordnet 中的超词。然而,Zeng 等人使用宽度为 3 的窗口层可以仅抽取特定大小的局部特征,却不能抽取更大范围的 n -gram 特征,并且他们在模型中加入了人工设定的词汇特征。

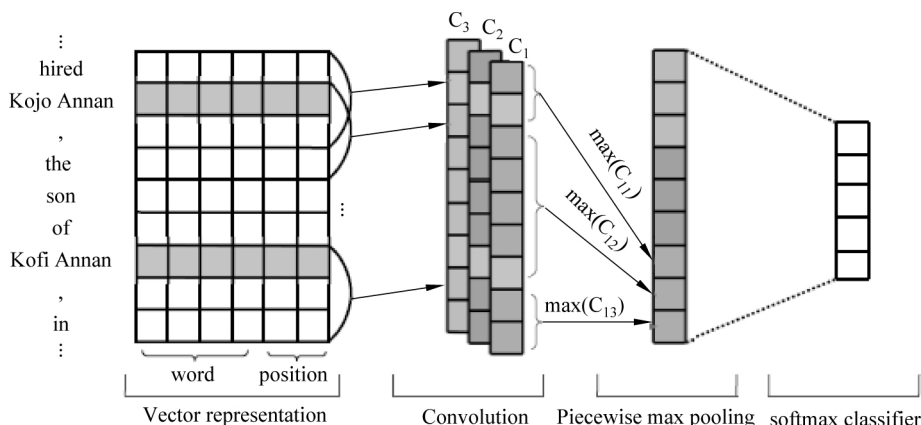


图 6 分段池化示意图^[46]

为了学习关系的更鲁棒表示,Xu 等^[9]在两个实体之间的最短依存路径上使用卷积神经网络学习关系表示,提出 depLCNN 模型,该模型的输入包含了带方向的箭头标记及关系类型。同时,Xu 等人将两个实体之间最短路径的逆向作为一个负关系来增加训练样本,使得模型显著提升了抽取效果。

在关系抽取任务中,每个单词对关系语义的贡献都是不等价的,因此注意力机制被自动地引入到神经网络中。Wang 等^[47]基于卷积神经网络,设计了输入注意力层和关系注意力层,并提出 BiAtt-pooling-CNN 模型。该方法使用捕获单词与目标实体相关性的输入注意力层和捕获单词与关系相关性的关系注意力层来获取句子特征向量表示,以尽可能突出句子中那些对关系抽取有更大贡献的部分。

BiAtt-pooling-CNN 模型仅仅使用输入注意力层即可达到 87.5% 的效果,再加上关系注意力层的效果则达到了 88.0%,这是当前研究中的最好结

针对这个问题,Nguyen 等^[45]使用多窗口联合抽取特征,以充分利用 n -gram 的特征并降低大窗口特征引起的噪音问题,同时完全不再使用词汇特征,仅仅使用端到端的神经网络学习句子的特征,取得了比 Zeng 等^[7]更好的效果。

以上方法使用最大池操作可以快速地减少隐层参数,但可能导致因此无法抽取句子的结构信息以及其他有效信息。

针对此问题,Zeng 等^[46]提出了 PCNN 模型,如图 6 所示,将卷积层得到的特征向量进行分段池化,即根据句子中两个实体的位置将卷积结果分为 3 段,对每段分别进行池化操作,将其组接作为当前句子的特征向量。使用分段池化操作可以保存句子部分结构信息,弥补句子太长对结果的影响。

果。从这里可以看出,相比关系注意力层,输入注意力层的作用更加重要。但是由于设计了关系注意力,使结构复杂度增长很多,且文章中设计了两个固定的权重矩阵,而事实上,不同的词在不同关系中的权重是不同的。

针对上述问题,Zhu 等^[48]提出 TCA-CNN 模型,仅使用捕获单词与目标实体相关性的输入注意力层,直接计算单词与关系的相关度,而不是计算单词与实体的相关性,同时去掉了关系注意力层。

相较于有监督方法而言,远程监督方法虽然可以快速得到大量有标注样本,然而不可避免地引入噪声。例如,包含“特朗普”和“美国”的句子可能还具有“特朗普出生于美国”的关系,或者这句话并不具有“特朗普是美国总统”关系。近年来,解决远程监督关系抽取的主要研究工作集中于在深度学习框架下的学习句包的表示,重点在于减少数据集中的噪声对关系抽取的影响。

Zeng 等^[46] 首先将深度学习模型与多示例学习结合起来,来进行远程监督关系抽取。Zeng 等使用 PCNN 模型获取句子的特征表示,在模型训练和预测阶段,仅选择每个句包中后验概率最大的句子作为句包的最终向量表示,来对模型参数进行更新。这个方法仅选择句包中最能表现句子中两个实体之间关系的句子,是个很强的限制,忽略了句包中大量的其他有用句子的信息。

针对“噪声问题”,Lin 等^[49] 在 Zeng 等^[46] 的基础上提出使用注意力机制给句包中的句子赋予不同的权重以降低噪声,因此可以更充分地利用信息。具体地,Lin 等^[49] 首先使用 CNN 对原始句子进行处理,得到句包中每个句子的向量表示;然后,使用注意力机制对每个句子计算一个权重,以表示在该关系上该句的重要程度;最后,使用该权重对句子向量表示进行加权求和,得到句包的向量表示,并将其送入到 softmax 中进行分类。每个句子根据关系的不同得到的权重都会不一样,这样就可以发现对于特定关系而言重要的句子与噪声的句子,并且利用句包中的所有信息预测关系类型。该方法中,关系的向量表示是随机初始化得到的。

为了更好地利用先验知识,Ji 等^[50] 在 Lin 等^[49] 的基础上,使用 PCNN 来获取句子的特征表示,同

时基于 Bordes 等^[51] 提出的 tranE 的思想,使用两个实体之间的差值作为关系的表示,并用来计算句子在句包中的权重,同时,引入 Freebase 和 Wikipedia 知识库里实体的文本描述信息来增强实体的语义表示。

卷积神经网络可以通过窗口层较好地学习句子中的局部特征,尤其是与关系相关的关键词信息。与序列结构的输入相比,引入树结构的输入可以减轻因为句子长度带来的噪声问题。

3.3 基于循环神经网络的关系抽取

由于句子含义跟单词出现的顺序是相关的,因此关系抽取可以看作是一个时序学习任务,可以使用循环神经网络来建模。

基于循环神经网络^[52] 的方法在模型设计上使用不同的循环神经网络来获取句子信息,然后对每个时刻的隐状态输出进行组合,在句子层级学习有效特征。在关系抽取问题中,对每一个输入,关系的标记一般只在序列的最后得到。

Zhang 等^[27] 首次使用双向循环神经网络来进行关系抽取,提出了 BRNN 模型。如图 7 所示,在双向循环神经网络中某一时刻的输出不仅依赖序列中之前的输入,也依赖于后续的输入。

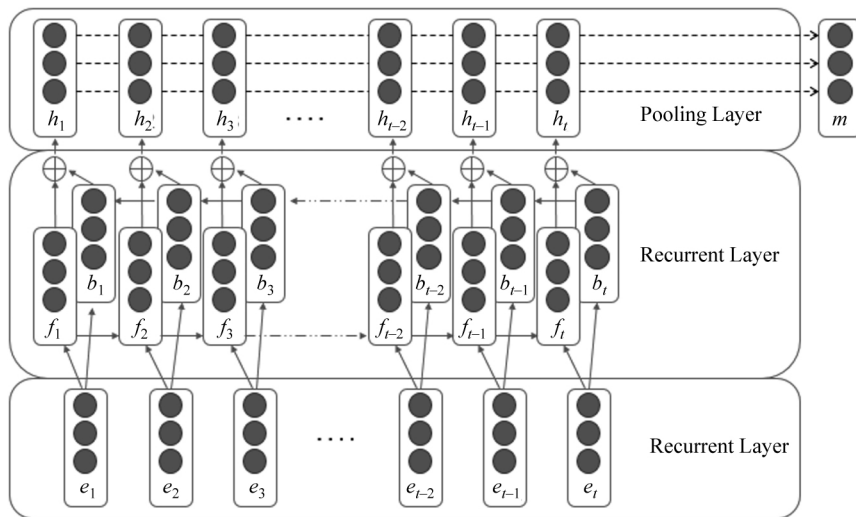


图 7 双向循环神经网络 BRNN 模型^[27]

对于给定的句子,Zhang 等^[27] 首先将每个单词按照正向和逆向两个顺序依次输入到两个方向不同的循环神经网络中,将正向和逆向隐状态加和作为当前时刻的隐状态输出。然后,对所有隐状态输出执行最大池操作,将得到的结果作为该句子的最终向量表示,并送入到一个逻辑斯蒂回归分类模型中。

这里使用最大池操作的出发点在于文章假设:只有几个重要的关键词(触发词)以及少部分的模式对关系抽取而言是重要的。

Hou 等^[30] 也使用双向循环神经网络模型进行关系抽取,他们认为关系的判定仅和实体附近的局部信息相关,因此其模型的输入是在两个实体的前

后固定长度的单词序列,并且给距离实体近的单词较大权重,给距离实体远的单词较小权重。

在简单循环神经网络中,“循环”指的是隐藏层中对每一个神经元的循环使用,隐藏层的输入不仅包括当前时刻的数据输入,还包括隐藏层自身上一时刻的输出^[53-54],会引起“梯度爆炸”问题。为了解决该问题,Hochreiter 等^[55]采用门机制,提出了长短期记忆神经网络(long short-term memory, LSTM)。

Xu 等^[37]首次使用 LSTM 网络捕捉句子信息来进行关系抽取。该文使用单词、词性、语法关系、Wordnet 超词等特征构建四个通道。在每个通道上,使用最短依存路径构造两个从实体到最小共同祖先节点的子序列,对每个子路径上的隐状态采取最大池化操作并串联组接,作为该句子在该通道的向量表示,最终将 4 个通道的结果组接起来,送入 softmax 函数中用来分类。

针对单向 LSTM 网络无法完全抽取上下文信息的问题,Zhang 等^[41]提出 BiLSTM 模型抽取句子双向的隐状态输出,并通过组合词汇特征和句子级别的两类特征来丰富句子的特征表示。其中,词汇特征有以下 7 个:单词本身、词性、实体类型、Wordnet 超词、位置特征、依存关系特征、相对依存位置特征。句子级别的特征通过对网络中两个实体的隐状态输出使用双重最大池化操作,加强两个实体之间的文本对关系抽取的影响。该文虽然取得了一定的效果,却仍然使用了大量的人工定义的特征。

Zhou 等^[56]在 Zhang 等^[27]的基础上,结合注意力机制提出了 Att-BiLSTM 模型。如图 8 所示,Zhou 等人使用双向 LSTM 网络来获取句子中每个单词对应的隐状态输出,使用注意力机制来抽取单词级别特征,通过特定的向量来计算每个单词对句子表示贡献的权重,以得到句子的最终向量表示用来分类。

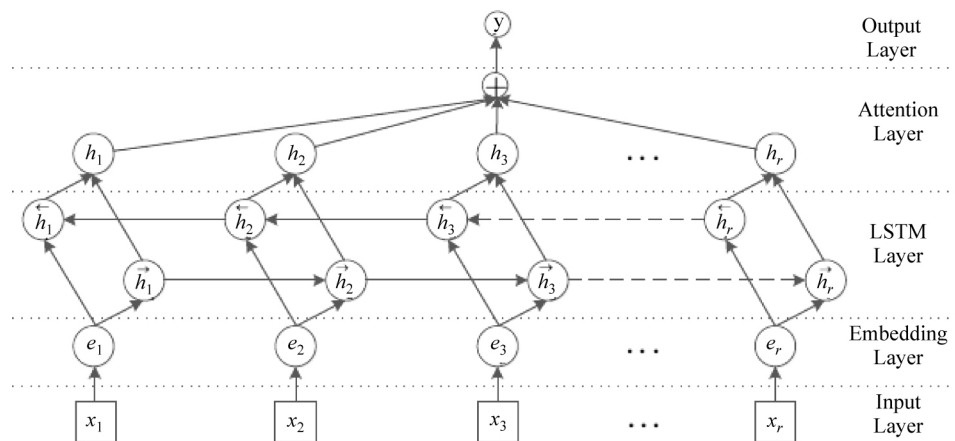


图 8 Zhou 等^[56]提出的 Att-BiLSTM 模型

事实上,实体对是关系抽取的基本成分,同一个单词与不同关系的权重是不同的,不同单词与实体对之间的顺序也反映了重要语义信息。在 Att-BiLSTM 模型中,注意力机制通过定义一个随机初始变量来表示当前单词与关系的相似度,这种设置并没有反映这些先验信息。

针对这个问题,Qin 等^[57]提出基于实体对的注意力机制,把(实体对,句子)作为输入,将实体对信息作为先验知识放入到注意力层,同时通过构建双向 GRU 网络^[58]以减少双向 LSTM 网络的计算复杂性。具体地,首先通过单向 GRU 网络得到实体的先验知识嵌入表示,以自适应地生成单词级别的注意力权重,网络的单向性也保证了这个嵌入表示与单词出现的顺序是相关的。随后,将注意力权

重与双向 GRU 网络上每个单词的隐状态输出整合,以获得句子的最终向量表示,并进行分类。

在 Qin 等人的基础上,Sun 等^[59]将注意力机制中的网络由 GRU 替换成 LSTM,然后使用粗粒度的双向 LSTM 网络来抽取句子级别特征信息,使用基于实体对注意力机制的细粒度双向 LSTM 来抽取关键信息。这种做法使得句子在包含很多噪声且关键信息无法正确抽取时,这两个粒度的网络功能可以互补。

一般认为,低层次网络能够进行足够的局部信息抽取,深层次网络可以探索不同层次和不同粒度的抽象特征。

为此,Xu 等^[10]在其自身工作^[37]的基础上,提出深层循环神经网络模型(DRNN)。具体地,在不

同特征构成的通道上,使用多层循环神经网络沿着两个子路径捕获抽象信息。深层网络有更强大的表达和学习能力,但是复杂性也提高了,同时需要更多的训练数据。Xu 等^[10]提出通过仅仅改变两个子路径的方向得到一个新的具有相反关系的样本的方法,以利用关系的方向性增加数据样本。

自注意力机制^[60]可以无视词之间的距离直接计算依赖关系,能够学习一个句子的内部结构,其中的多头注意力可以允许模型在不同的表示子空间里学习到相关的信息的特点引起研究人员关注。为此, Lee 等^[61]使用多头注意力对句子的每个单词的上下文信息及隐实体类型信息进行编码,随后使用双向 LSTM 网络对句子进行编码,并叠加词语级别的注意力机制来获取句子的向量表示。

上述提到的方法在实际上仅考虑关系分类子过程的工作,未能将两个子过程联合考虑,属于流水线学习方法范畴。为了解决该问题, Miwa 等^[62]在双向序列 LSTM 的基础上堆叠双向树结构的 LSTM 进行联合关系抽取。具体地, Miwa 等人先使用双向序列 LSTM 来识别所有实体,然后再遍历所有的实体对,使用树结构双向 LSTM 网络来输出它们之间的关系。

该方法在一个模型框架下进行实体识别和关系分类,两个子任务通过共享网络参数的方式进行联合学习。但是在训练时,这两个子过程是有先后关系的,需要先进行实体识别,再根据实体识别的结果进行两两匹配来进一步进行关系抽取。

为了解决上述问题, Zheng 等^[11]采取联合学习的方法,提出一种新的标注策略来进行关系抽取,将关系抽取转换成了序列标注问题。Zheng 等人使用双向 LSTM 网络对句子进行编码,然后使用 LSTM 网络对输入单词的隐状态进行解码。进行标注的时候,每个输入节点的标注由三个部分组成,分别是:单词在实体中的位置;关系类型;关系角色(即关系方向)。最后,根据标签序列,将同样关系类型的实体合并成一个三元组作为最后抽取结果。这种通过序列标注的方法可以同时得到对应的实体类型和实体对之间的关系,两个子过程同时进行。

基于循环神经网络的方法通过合理地引入神经网络的各类机制,较好地学习句子中的特征,在对长距离关系建模时具有天然优势。

3.4 基于混合网络模型的关系抽取

为了更好地抽取句子中的特征,研究人员使用

递归神经网络、卷积神经网络与循环神经网络 3 种网络及其他机器学习方法进行组合建模来进行关系抽取。

为了更好地利用依存树上的信息, Liu 等^[39]构造了基于增强依存子树的网络模型 DepNN。DepNN 首先使用递归神经网络来对增强依存子树上每个子树进行建模,获取其子树向量表示,然后使用卷积神经网络来捕获最短依存路径上的重要的局部特征,作为句子的最终向量表示进行关系抽取。

Vu 等^[63]提出了基于文本扩展表示的 ECNN 和基于链接的 UniBRNN 模型,将每个神经网络得到的多个结果根据投票机制得到关系的最终抽取结果。文本扩展表示的思想是:大多数情况下,两个实体中间的子句序列包含与实体抽取最相关的单词,我们主要关注这个子序列,同时不能忽略其他子序列。ECNN 首先用输入的句子构造两个子序列,第一个子序列为第二个实体之前的子句,第二个子序列为第一个实体之后的子句,然后在 Liu 等^[33]的基础上分别对两个子序列进行处理,将其结果串联得到句子的向量表示。UniBRNN 区别于传统的 BRNN 的地方在于:对于一个输入序列,特定时刻的隐状态输出与逆序列相同时刻的隐状态进行串联组合,而不是序列在当前节点的逆向隐状态输出,并且对于中间隐节点也进行前向 RNN 操作。

Xiao 等^[54]将注意力机制引入一个多级的循环神经网络,该方法使用文本序列作为输入,根据标记实体的位置将句子分为 5 部分,使用同一个双向 LSTM 网络在 3 个子序列上独立学习,然后引入词层级的注意力机制关注重要的单词表示,分别得到子序列的向量表示;随后,使用双向 RNN 网络进一步抽取子序列和实体的特征,并再次使用注意力机制将其转换成句子的最终向量表示,并送入到分类器中。Nguyen 等^[64]将传统基于特征的方法(log-linear 模型)、卷积神经网络方法和循环神经网络方法使用集成、投票等机制进行组合。Zhang 等^[65]提出在双向 LSTM 的基础上叠加注意力机制,以及使用卷积神经网络层获取句子的表示,再送入到一个全连接层和 softmax 层进行分类。

在包含两个实体的最短依存路径上,先前的工作多是把依存关系当作单词或者词性等语法特征,忽略了特征之间的差异。针对这个问题, Cai 等^[66]提出双向 RCNN 模型记为(BRCNN),通过在双通道 LSTM 网络上叠加卷积神经网络,来利用最短依存树上的依存关系。具体地,在每个方向的循环神

神经网络中,将整个最短依存树中的单词和单词之间的关系使用不同类的嵌入表示,并且将单词与关系分别使用独立通道的 LSTM 编码。在卷积层,将单向循环神经网络层获取的相邻词对应的 LSTM 输出和它们的依存关系层的 LSTM 输出连结起来作为卷积层的输入,在卷积层后接池化层,在池化层后接 softmax 分类。

从 Wikipedia、Dbpedia、Wikidata 等百科全书网站上有很多关于实体的描述,这些描述可以为实体提供额外的描述信息,提升分类效果。针对这个事实,Ren 等^[67]在 Cai 等^[66]的基础上提出 DesRC 模型。模型的输入为(句子,实体 1 的描述,实体 2 的描述),然后分别使用卷积神经网络抽取实体的描述信息,使用 BRCNN 模型抽取句子的向量表示,最后使用跨句子的注意力方法来将其整合成一个全局实值变量,随后将其送入到 softmax 层进行分类。

为了更有效提升依存模型的计算效率,利用最短依存路径有可能忽略的重要的信息,Zhang 等^[68]对依存解析树使用 path-centric 的剪枝,即保留依存树上和最短依存路径上的每个节点邻接的数个节点,然后使用双向 LSTM 网络学习依存树上节点的上下文信息,并在此基础上设计了面向关系抽取的图卷积神经网络,通过池化操作来获取剪枝的依存树的句子表示,将其与两个经过图卷积的实体的向量表示组合,作为句子的最终向量表示来进行关系抽取。

在联合学习问题上,Zheng 等^[69]使用递归神经网络和卷积神经网络组合来进行联合学习,也是一种共享底层网络参数的方法。具体地,对于输入句子使用双向 LSTM 网络来进行编码,然后用一个 LSTM 网络来进行实体抽取,用一个卷积神经网络来进行关系抽取。接下来将实体抽取的预测标签进行嵌入表示,再传入到实体抽取的解码中来代替条件随机场解决命名实体识别中的标签依赖问题。在进行关系抽取的时候,需要先根据命名实体识别预测的结果对实体进行配对,然后将实体之间的文本使用卷积神经网络进行关系抽取。

在远程监督问题上,针对噪声问题,Chen 等^[70]使用 PCNN 来学习句子的表示,然后使用多层感知机和逻辑回归来捕捉噪声的信息,过滤掉信息量较小的句子,仅选择句包中最重要的数个句子,随后使用注意力机制对选定的句子以不同权重来计算句包的向量表示,来减少噪声残留问题,最后再将其送入到 softmax 层中进行分类。

之前的方法,多是针对一个句子只有一个关系的情况进行处理。实际上,一个句子可能存在多个关系。同时,之前采取的方法都是从句子级别的特征抽取,而不是跨句子的特征抽取,而事实上两个实体之间的语义关系可能由多个句子组合来表达。

针对上述问题,Jiang 和 Wang 等人分别提出多示例多标签 CNN 模型^[71]和 DSNN 模型^[72]来进行关系抽取。具体地,两人分别使用 CNN 和 PCNN 获取句子的向量表示,同时基于张量可以更好地描述特征之间的相关性的特点,Wang 还采取张量转移的方法获取句子更为最丰富的特征表示。然后二者均跳过句子级别的抽取,直接对句包内部的所有句子的特征表示做最大池化操作。在训练阶段,对于每个关系,使用训练集中所有句包来训练一个逻辑回归分类器,在预测阶段,每个句包就可能对应多个关系。

为了更有效学习句包的特征表示,Feng 等^[73]使用两个内存网络来分别计算每个单词与不同的实体对的相关性以及不同关系之间的依赖性,以此来挖掘深层次的关系。第一个内存网络实现了单词级别的注意力机制,即根据上下文与实体对的相似度,对单词分配权重,将其与通过 CNN 网络获取的句子特征表示进行组接,代表句子的最终表示。第二个内存网络实现了句子级别和句包级别的两段注意力机制,第一段注意力与 Lin 等^[49]的方法类似,通过计算当前句包内的句子与关系的权重,得到当前句包在不同关系下的向量表示;第二段注意力通过分别计算当前关系与其他关系之间的相关性权重,对每个句包在不同关系下的表示进行加权求和,就可以得到融合多关系的句包的特征向量。对于最后得到的每一个句包的特征向量,使用逻辑回归做多个二元分类器,得到句包的多个可能关系。

上述方法中,句包的标签是固定不变的,而事实上,随着学习的进行,由于特定样本与正确标记示例的句法或者语义相似性,样本被正确分类的可能性在提升,某些被错误分类的句子也可能被正确划分。

针对这个现象,Liu 等^[74]提出软标签方法,该方法在训练过程中句包的标签是动态变化的,可以在训练阶段利用这种语义或句法相似性来纠正误标记样本。在训练过程中,Liu 等人提出一个联合得分函数来获取软标记样本,这个函数既考虑样本的置信度,也考虑句包的表示。在得到句包的表示以后,由于使用软标签,可以根据训练的关系矩阵计算当前句包属于每个关系的权重,最后使用一个置信向

量来表示当前标签的置信度,并将最可能的关系进行放大。

针对噪声问题,Feng^[75]、Zeng^[76]和 Qin^[77]等均使用 CNN 网络抽取句子的特征标识,然后使用强化学习或生成对抗训练的策略来减少训练数据集中的假正样本。

Feng 等^[75]的模型有两个部件,分别是实例选择器和关系分类器,实例选择器对每个句子执行动作,选择是否将句子放入训练集;分类器对样本进行预测,并将预测结果回传给实例选择器,帮助实例选择器更好地选择训练样本。Zeng 等^[76]将“至少一次”假设进行了重新表述:在对句包进行关系预测时,当且仅当每个句子的预测标签都是“无关系”时,句包才是“无关系”,否则句包的关系类型由其中的句子所表达,从另一个角度使用了实例选择器。Qin 等^[77]通过在训练数据集上训练生成器,使得生成器生成的数据被判别器标识成负样本,最终使用生成器过滤掉有噪声样本的方法来去除数据集中的噪声。

基于混合网络模型的关系抽取工作重点在于结合不同的语言学知识,使用不同的神经网络的结构特点来共同设计网络模型以达到理想的效果。

3.5 小结

本节主要从模型构建的角度对基于深度学习的关系抽取方法进行了分析和总结,可以发现以下几点。

(1) 这几种深度学习的模型在关系抽取任务上的区别主要其处理 NLP 领域任务中的作用不同:递归神经网络考虑的是获得符合语法规则的短语特

征,卷积神经网络考虑的是所有可能的短语组合,循环神经网络在根据当前词的上下文来获得有效特征方面具有独特的优势。通过模型的组合也可以一定程度上提升关系抽取的效果,并且注意力机制已经成为关系抽取领域的常见做法。

(2) 从形式上来讲,基于递归神经网络的关系抽取方法要求根据输入构造特定的树结构,然而相同节点之间的组合方式过于简单,捕捉语义的作用不是很明显。由于解析树的构造比较复杂,而且在递归神经网络中进行简单的组合也难以足够抽取句子的语义关系特征,给每个单词赋予一个向量或矩阵也导致计算量太过庞大,现已较少使用。卷积神经网络由于训练过程中参数较少,效果优异,成为在关系抽取领域模型构建中获取句子特征表示时的常用模型。

(3) 在基于深度学习的关系抽取应用早期,使用序列结构的输入是卷积神经网络和循环神经网络的常用做法。近年来,随着研究的深入,图结构的输入也引起了研究人员的关注。尽管构建树结构的输入需要利用其他自然语言处理工具,导致增加计算量,并存在误差传递问题,但由于其在捕获语义方面的独特优势,使用树结构的输入仍然作为基于深度学习的关系抽取的重要形式。

(4) 这里将几个经典的基于深度学习的关系抽取模型在特定数据集上得到的结果进行对比,如表 1 所示。可以发现,为了获得句子的向量表示,构建网络模型时仍需考虑如何正确反映一些自然语言处理方面的先验知识,比如 n -gram 特征、依存关系特征、句法成分特征、单词上下文特征等各类自然语言特征。

表 1 有监督关系抽取方法比较

模型	F_1 值	特征	其他技术
MV-RNN ^[8]	82.4	单词、词性、语法关系、Wordnet 超词	
CNN ^[35]	82.7	单词、单词位置、Wordnet 超词	
Chain-CNN ^[36]	82.7	单词、词性、语法关系、Wordnet 超词	
SDP-LSTM ^[7]	83.7	单词、词性、语法关系、Wordnet 超词	
Att-BiLSTM ^[56]	84.0	单词、实体标识	
CR-CNN ^[28]	84.1	单词、单词位置	基于 margin 的 Ranking loss
depLCNN ^[9]	85.6	单词、实体周边单词、Wordnet 超词	负采样技术
DRNN ^[10]	86.1	单词、词性、语法关系、Wordnet 超词	数据增强
TCA-CNN ^[48]	87.3	单词、单词位置	基于 margin 的成对损失函数
BiAtt-pooling-CNN ^[47]	88.0	单词、单词位置	基于 margin 的 Ranking loss

另外,通过合理利用训练样本数据标记特征设计特定的损失函数,如基于 margin 的 ranking loss^[28],Zhu 等^[48]提出基于 margin 的成对损失函数,代替原来的交叉熵损失也可以提升抽取效果。其中 margin 的设定使得模型尽可能让不同类别样本在样本空间中距离更远。通过负采样技术和样本增强技术可以高效增加样本数量,一定程度上提升抽取效果。

(5) 不论是有监督关系抽取还是远程监督关系抽取,从一个句子或者特定的语料中都很难获得句中实体足够的先验信息。采取融入外部知识或者已有知识库的结构化知识等方法,如通过 Wordnet 超词获取更加丰富的语义信息,以及从相关知识库获取实体文本描述信息和关系信息的方式引入领域常识,已成为提升抽取效果的有效做法。

4 未来的研究重点

近年来,深度学习作为关系抽取的主要技术,得到了很好的发展,然而在具体应用过程中还有较长的路要走。本节对使用深度学习的关系抽取的未来发展方向进行展望。

4.1 结合深度学习进展的关系抽取研究

在深度学习被广泛使用后,其抽取效果也依赖于神经网络的研究发展。2014 年至 2016 年,该领域的研究热点集中于递归神经网络、卷积神经网络、循环神经网络等基本网络结构的研究。2017 至今,随着 Attention 机制^[61]、强化学习^[75]、生成对抗学习^[72]、图神经网络^[78]以及预训练技术^[79]等方法的研究深入,又有各类不同的网络被设计用来抽取关系。在基于深度学习的关系抽取框架下,如何结合各类技术进行关系抽取的研究是未来研究的一个重要方向。

4.2 面向关系抽取的模型设计与研究

现有的关系抽取方法已经覆盖了多数可能的输入和特征,如何在引入领域知识的同时直接构造端到端的抽取模型,不再使用各类自然语言处理工具,以减少误差的传递,仍然值得探索。

基于深度学习的模型往往依赖于网络的复杂程度带来的对隐含信息挖掘能力的提升。现有网络在设计时考虑的多是语言学直观特性,在理论可解释性方面有待开展更深层次的研究。

4.3 面向关系抽取的迁移学习

在关系抽取任务中,常用的关系抽取数据集面临着关系类别和样本数量较少的问题。事实上,在命名实体识别、文本分类等自然语言处理任务中,同样有着大量的有标注数据集,而这些任务在语义级层面有着一定的相关性。因此,除了构建新的有标注训练数据集方法以外,在深度学习框架下,如何有效利用与关系抽取任务相关的数据集或模型,将已有的数据集及模型迁移到关系抽取任务中来,也值得期待。

4.4 篇章级关系抽取

仅仅从给定的实体和句子中判断其关系往往缺少必要的语义知识,直观上来讲人们在具备一定的知识的基础上,阅读了整段文字或者整篇文章可能对实体之间的关系才能了解得更加深刻。因此,如何结合已有的知识信息进行实体与关系的联合抽取以及进行跨句子、跨段落等更粗粒度的关系抽取,具有更实际的研究价值。

4.5 多元关系抽取

以往的文献多集中于单个句子的二元关系抽取,然而在实际应用中,我们面对的数据中存在着多个实体以及复杂关系,比如像事件类型的数据,可能涉及事件名称、时间、地点等层次语义信息,二元关系抽取无法一次性抽取上述论元,难以满足任务需求。为此,有待开展更深层次的研究,进行多元关系抽取。

4.6 面向开放域的关系抽取

本文提到的常用的数据集的关系类别和训练数据都是固定的,因此基于深度学习的关系抽取多属于预定义的关系抽取。然而,仅仅在 Freebase 数据库中,已有 4 000 多万个实体,24 亿多个事实三元组,上万个属性关系。互联网上每天都产生海量的跨领域的复杂数据,而且这些数据往往包含着多种关系,封闭域的二元关系抽取很难满足网络文本抽取的需求。如何结合弱监督和半监督的知识,使用深度学习方法进行开放域的关系抽取,具有重要的应用价值。

5 结论

关系抽取作为信息抽取技术的一环,一直以来

受到了相当广泛的关注。深度学习在自然语言处理领域的重要突破,也受到了人们的重视,被引入到关系抽取领域中来。本文详细阐述了基于深度学习的一些关系抽取方法,对相关工作难点和挑战进行了总结。并在此基础上,对未来的研究方向进行了分析和展望。

参考文献

- [1] Wu F, Weld D S. Open information extraction using Wikipedia[C]//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2010: 118-127.
- [2] Fleischman M, Hovy E, Echihi A. Offline strategies for online question answering: Answering questions before they are asked[C]//Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1. Association for Computational Linguistics, 2003: 1-7.
- [3] Goodfellow I, Bengio Y, Courville A. Deep learning [M]. MIT press, 2016.
- [4] Plank B, Søgaard A, Goldberg Y. Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss[J]. arXiv preprint arXiv: 1604.05529, 2016.
- [5] Lee J Y, Dernoncourt F, Szolovits P. Transfer learning for named-entity recognition with neural networks [J]. arXiv preprint arXiv: 1705.06273, 2017.
- [6] Zhou J, Xu W. End-to-end learning of semantic role labeling using recurrent neural networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, 1: 1127-1137.
- [7] Zeng D, Liu K, Lai S, et al. Relation classification via convolutional deep neural network[C]//Proceedings of the COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 2014: 2335-2344.
- [8] Socher R, Huval B, Manning C D, et al. Semantic compositionality through recursive matrix-vector spaces[C]//Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics, 2012: 1201-1211.
- [9] Xu K, Feng Y, Huang S, et al. Semantic relation classification via convolutional neural networks with simple negative sampling [J]. arXiv preprint arXiv: 1506.07650, 2015.
- [10] Xu Y, Jia R, Mou L, et al. Improved relation classification by deep recurrent neural networks with data augmentation[J]. arXiv preprint arXiv: 1601.03651, 2016.
- [11] Zheng S, Wang F, Bao H, et al. Joint extraction of entities and relations based on a novel tagging scheme [J]. arXiv preprint arXiv: 1706.05075, 2017.
- [12] Kumar S. A survey of deep learning methods for relation extraction [J]. arXiv preprint arXiv: 1705.03645, 2017.
- [13] 徐建, 张智雄, 吴振新, 实体关系抽取的技术方法综述 [J]. 现代图书情报技术, 2008, 168(8): 18-23.
- [14] 刘方驰, 钟志农, 雷霆, 等. 基于机器学习的实体关系抽取方法[J]. 兵工自动化, 2013, 32(9): 57-62.
- [15] Xun H, Hongliang Y, Yang Y. A review of relation extraction[J]. Data Analysis and Knowledge Discovery, 2013, 29(11): 30-39.
- [16] Mintz M, Bills S, Snow R, et al. Distant supervision for relation extraction without labeled data[C]//Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2. Association for Computational Linguistics, 2009: 1003-1011.
- [17] Brin S. Extracting patterns and relations from the world wide web[C]//Proceedings of the International Workshop on The World Wide Web and Databases. Springer, Berlin, Heidelberg, 1998: 172-183.
- [18] Hasegawa T, Sekine S, Grishman R. Discovering relations among named entities from large corpora[C]//Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004: 415.
- [19] Zhao S, Grishman R. Extracting relations with integrated information using kernel methods[C]//Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2005: 419-426.
- [20] Kambhatla N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations[C]//Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions. Association for Computational Linguistics, 2004: 22.
- [21] Sam R C, Le H T, Nguyen T T. Relation extraction in Vietnamese text using conditional random fields [C]//Proceedings of Asia Information Retrieval Symposium. Springer, Berlin, Heidelberg, 2010: 330-339.
- [22] Zelenko D, Aone C, Richardella A. Kernel methods for relation extraction[J]. Journal of Machine Learning Research, 2003, 3(Feb): 1083-1106.

- [23] Katiyar A, Cardie C. Going out on a limb: Joint extraction of entity mentions and relations without dependency trees[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017, (1): 917-928.
- [24] Doddington G R, Mitchell A, Przybocki M A, et al. The automatic content extraction (ACE) program-tasks, data, and evaluation[C]//Proceedings of the 4th International Conference on Language Resources and Evaluation, 2004: 837-840.
- [25] Hendrickx I, Kim S N, Kozareva Z, et al. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals[C]//Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions. Association for Computational Linguistics, 2009: 94-99.
- [26] Riedel S, Yao L, McCallum A. Modeling relations and their mentions without labeled text[C]//Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2010: 148-163.
- [27] Zhang D, Wang D. Relation classification via recurrent neural network[J]. arXiv preprint arXiv: 1508.01006, 2015.
- [28] Santos C N, Xiang B, Zhou B. Classifying relations by ranking with convolutional neural networks[J]. arXiv preprint arXiv: 1504.06580, 2015.
- [29] Rink B, Harabagiu S. Utd: Classifying semantic relations by combining lexical and semantic resources [C]//Proceedings of the 5th International Workshop on Semantic Evaluation. Association for Computational Linguistics, 2010: 256-259.
- [30] Hou X, Zhao Z, Liu T, et al. Classifying relation via bidirectional recurrent neural network based on local information[C]//Proceedings of the Asia-Pacific Web Conference. Springer, Cham, 2016: 420-430.
- [31] Zheng S, Xu J, Zhou P, et al. A neural network framework for relation extraction: Learning entity semantic and relation pattern [J]. Knowledge-Based Systems, 2016, 114: 12-23.
- [32] Qin P, Xu W, Guo J. An empirical convolutional neural network approach for semantic relation classification[J]. Neurocomputing, 2016, 190: 1-9.
- [33] Liu C Y, Sun W B, Chao W H, et al. Convolution neural network for relation extraction[C] //Proceedings of the International Conference on Advanced Data Mining and Applications. Springer, Berlin, Heidelberg, 2013: 231-242.
- [34] Hashimoto K, Miwa M, Tsuruoka Y, et al. Simple customization of recursive neural networks for semantic relation classification[C]//Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 2013: 1372-1376.
- [35] Yang Y, Tong Y, Ma S, et al. A position encoding convolutional neural network based on dependency tree for relation classification[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016: 65-74.
- [36] Ebrahimi J, Dou D. Chain based RNN for relation classification[C]//Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015: 1244-1249.
- [37] Xu Y, Mou L, Li G, et al. Classifying relations via long short term memory networks along shortest dependency paths[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 1785-1794.
- [38] Nguyen D, Matsuo Y, Ishizuka M. Relation extraction from wikipedia using subtree mining[C]//Proceedings of the 22nd National Conference on Artificial Intelligence-Volume 2. AAAI Press, 2007: 1414-1420.
- [39] Liu Y, Wei F, Li S, et al. A dependency-based neural network for relation classification[J]. arXiv preprint arXiv: 1507.04646, 2015.
- [40] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch[J]. Journal of Machine Learning Research, 2011, 12 (Aug): 2493-2537.
- [41] Zhang S, Zheng D, Hu X, et al. Bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation. 2015: 73-78.
- [42] Frege G. Über sinn und bedeutung[J]. Zeitschrift für Philosophie und philosophische Kritik, 1892, 100 (1): 25-50.
- [43] Mnih V, Heess N, Graves A. Recurrent models of visual attention[C]//Proceedings of the Advances in Neural Information Processing Systems, 2014: 2204-2212.
- [44] Socher R, Pennington J, Huang E H, et al. Semi-supervised recursive autoencoders for predicting sentiment distributions[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011: 151-161.
- [45] Nguyen T H, Grishman R. Relation extraction: Perspective from convolutional neural networks [C]//Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, 2015:

- 39-48.
- [46] Zeng D, Liu K, Chen Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 1753-1762.
- [47] Wang L, Cao Z, De Melo G, et al. Relation classification via multi-level attention cnns[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 1298-1307.
- [48] Zhu J, Qiao J, Dai X, et al. Relation classification via target-concentrated attention cnns[C]//Proceedings of the International Conference on Neural Information Processing. Springer, Cham, 2017: 137-146.
- [49] Lin Y, Shen S, Liu Z, et al. Neural relation extraction with selective attention over instances[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016, (1): 2124-2133.
- [50] Ji G, Liu K, He S, et al. Distant supervision for relation extraction with sentence-level attention and entity descriptions[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence, 2017.
- [51] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data [C]//Proceedings of the Advances in Neural Information Processing Systems, 2013: 2787-2795.
- [52] 邱锡鹏. 神经网络与深度学习[EB/OL]. [2017-04-21]. <https://nndl.github.io/>.
- [53] Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model[C]//Proceedings of the 11th Annual Conference of the International Speech Communication Association, 2010.
- [54] Xiao M, Liu C. Semantic relation classification via hierarchical recurrent neural network with attention [C]//Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics; Technical Papers, 2016: 1254-1263.
- [55] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [56] Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016, (2): 207-212.
- [57] Qin P, Xu W, Guo J. Designing an adaptive attention mechanism for relation classification [C]//Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, 2017: 4356-4362.
- [58] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv: 1406.1078, 2014.
- [59] Sun Y, Cui Y, Hu J, et al. Relation classification using coarse and fine-grained networks with SDP supervised key words selection[C]//Proceedings of the International Conference on Knowledge Science, Engineering and Management. Springer, Cham, 2018: 514-522.
- [60] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the Advances in Neural Information Processing Systems, 2017: 5998-6008.
- [61] Lee J, Seo S, Choi Y S. Semantic relation classification via bidirectional LSTM networks with entity-aware attention using latent entity typing[J]. arXiv preprint arXiv: 1901.08163, 2019.
- [62] Miwa M, Bansal M. End-to-end relation extraction using LSTMs on sequences and tree structures[J]. arXiv preprint arXiv: 1601.00770, 2016.
- [63] Vu N T, Adel H, Gupta P, et al. Combining recurrent and convolutional neural networks for relation classification[J]. arXiv preprint arXiv: 1605.07333, 2016.
- [64] Nguyen T H, Grishman R. Combining neural networks and log-linear models to improve relation extraction [J]. arXiv preprint arXiv: 1511.05926, 2015.
- [65] Zhang X, Chen F, Huang R. A combination of RNN and CNN for attention-based relation classification [J]. Procedia Computer Science, 2018, 131: 911-917.
- [66] Cai R, Zhang X, Wang H. Bidirectional recurrent convolutional neural network for relation classification [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016, (1): 756-765.
- [67] Ren F, Zhou D, Liu Z, et al. Neural relation classification with text descriptions[C]//Proceedings of the 27th International Conference on Computational Linguistics, 2018: 1167-1177.
- [68] Zhang Y, Qi P, Manning C D. Graph convolution over pruned dependency trees improves relation extraction[J]. arXiv preprint arXiv: 1809.10185, 2018.
- [69] Zheng S, Hao Y, Lu D, et al. Joint entity and relation extraction based on a hybrid neural network[J]. Neurocomputing, 2017, 257: 59-66.
- [70] Chen Y, Liu H, Wu C, et al. Distant supervision for relation extraction with neural instance selector[C]// Proceedings of the CCF International Conference on

- Natural Language Processing and Chinese Computing. Springer, Cham, 2018; 209-220.
- [71] Jiang X, Wang Q, Li P, et al. Relation extraction with multi-instance multi-label convolutional neural networks[C]//Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016; 1471-1480.
- [72] Wang Z, Chang B, Sui Z. Distantly supervised neural network model for relation extraction[M]. Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Springer, Cham, 2015; 253-266.
- [73] Feng X, Guo J, Qin B, et al. Effective deep memory networks for distant supervised relation extraction [C]//Proceedings of the IJCAI, 2017; 4002-4008.
- [74] Liu T, Wang K, Chang B, et al. A soft-label method for noise-tolerant distantly supervised relation extraction[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017; 1790-1795.
- [75] Feng J, Huang M, Zhao L, et al. Reinforcement learning for relation classification from noisy data [C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018.
- [76] Zeng X, He S, Liu K, et al. Large scaled relation extraction with reinforcement learning[C] //Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018.
- [77] Qin P, Xu W, Wang W Y. Dsgan: Generative adversarial training for distant supervision relation extraction[J]. arXiv preprint arXiv: 1805.09929, 2018.
- [78] Wang S, Zhang Y, Che W, et al. Joint extraction of entities and relations based on a novel graph scheme [C]//Proceedings of the IJCAI, 2018; 4461-4467.
- [79] Wu S, He Y. Enriching pre-trained language model with entity information for relation classification[J]. arXiv preprint arXiv: 1905.08284, 2019.



庄传志(1985—),博士研究生,工程师,主要研究领域为机器学习与数据挖掘。

E-mail: zhuangcz@ict.ac.cn



朱伟建(1980—),博士研究生,主要研究领域为计算机取证、数据挖掘。

E-mail: zhuweijian@ict.ac.cn



靳小龙(1976—),博士,研究员,主要研究领域为大数据知识工程、知识计算、知识图谱。

E-mail: jinxiaolong@ict.ac.cn