

文章编号: 1003-0077(2020)03-0072-08

基于远程监督的藏文实体关系抽取

王丽客^{1,2}, 孙媛^{1,2}, 夏天赐^{1,2}

(1. 中央民族大学 信息工程学院, 北京 100081;

2. 中央民族大学 国家语言资源监测与研究中心 少数民族语言分中心, 北京 100081)

摘要: 关系抽取任务是对句子中的实体对进行关系分类。基于远程监督的关系抽取是用预先构建的知识库来对齐朴素文本, 自动标注数据, 在一定程度上减少了人工标注的成本, 缓解了藏文材料语料不足的问题。但是基于远程监督的实体关系抽取还存在错误标记、提取特征时出现噪声等问题。该文用远程监督方法进行藏文实体关系抽取, 基于已经构建的藏文知识库, 利用分段卷积神经网络结构, 加入语言模型和注意力机制来改善语义歧义问题以及学习句子的信息; 在训练过程中加入联合得分函数来动态修正错误标签问题。实验结果表明改进的模型有效提高了藏文实体关系抽取的准确率, 且优于基线模型效果。

关键词: 藏文实体关系抽取; 语言模型; 注意力机制

中图分类号: TP391

文献标识码: A

Distant Supervision for Tibetan Entity Relation Extraction

WANG Like^{1,2}, SUN Yuan^{1,2}, XIA Tianci^{1,2}

(1. School of Information Engineering, Minzu University of China, Beijing 100081, China;

2. Minority Languages Branch, National Language Resource and Monitoring Research Center, Minzu University of China, Beijing 100081, China)

Abstract: Distant supervision for relation extraction is an efficient method to automatically align entities in texts to a given knowledge base (KB), which alleviated the problem of manual labelling. In this paper, we propose an improved distant supervised relation extraction model in Tibetan based on Piecewise Convolutional Neural Network (PCNN). The language model and the selective-attention mechanism are combined to alleviate wrong labelling problems and to extract effective features. Soft-label method is also introduced to dynamically correct the relation label. The experimental results show that our method is effective and outperforms several competitive baseline methods.

Keywords: Tibetan entity relation extraction; language model; attention mechanism

0 引言

实体关系抽取任务作为信息抽取领域的重要研究课题, 从无结构化文本中抽取句子中已标记的实体对之间的语义关系, 可以应用于知识图谱和问答系统的构建。实体关系抽取可以使用无监督、半监督、完全监督和远程监督的方法, 无监督和半监督方法使用 OpenIE 工具, 没有预定义本体和关系类别, 直接从数据和关系短语中提取事实。有监督的方法需要大量的训练数据, 由于完全监督方法要求完全

正确地标注数据集, 因而数据量很小, 为了避免人工构建用于关系抽取的数据集, Mintz 等^[1]提出远程监督的方法, 通过对齐知识库与文本, 来自动生成大规模的训练数据, 预测文本中实体对之间的语义关系, 将这种对齐用来训练关系抽取器。本文首次尝试将远程监督的方法用在藏文实体关系抽取任务上, 即对于知识库中的三元组 $\langle e_1, r, e_2 \rangle$, 所有在文本中提到实体 e_1 和 e_2 的句子都被认为是关系 r 的训练数据。

就实体关系抽取任务来说, 中文和英文的使用范围非常广泛, 方法和模型也很先进, 相比而言, 藏

收稿日期: 2019-09-16 定稿日期: 2019-10-16

基金项目: 国家自然科学基金(61972436)

文是少数民族语言,认识和传播的力度不够,藏文母语人士缺乏,而且对藏文的研究也处于起步阶段,导致人工标注藏文训练语料费时费力。远程监督方法刚好弥补这一缺陷,可以自动标注训练语料。例如,知识库中有实体对<ཇོ་བོ་(乔布斯),འཕྲི(苹果)>,在语料库中包含该实体对的句子可能多个,例如,ཇོ་བོ་ནི་ཀུན་གྱི་གཞི་རྒྱུ་ལས་འཛིན་པ་ཡིན།(乔布斯是苹果公司的总裁。)和ཇོ་བོ་ཀུན་གྱི་ཐུ་རྩ་དགུ།(乔布斯喜欢吃苹果)。远程监督方法会将所有包含该三元组<ཇོ་བོ་(乔布斯),/people/company/founders,འཕྲི(苹果)>的句子都提取出来作为一个包并作为该关系的训练语句。在上述描述中,第1句确实表示/people/company/founders的关系,而第2句中的两个实体只是同时出现,并没有表示该关系。这说明远程监督方法的假设性太强,会导致错误标记问题。此外,在训练语料中还存在一词多义的现象,如上述“འཕྲི(苹果)”一词,需要根据上下文语境才能判断取“食物苹果”还是“苹果公司”的含义。

首先,本文加入 ELMo 语言模型,根据输入句子的上下文语境,动态生成词向量,来解决语义歧义问题。其次,为了更有效地利用句子的信息以及自动获取句子特征,在分段卷积神经网络结构中加入选择性注意力机制,来自动学习句子特征,并为包内的每个句子分配权重。最后,为了减缓错误标记问题,本文将预测的关系标签与原始远程监督生成的标签进行联合评价,动态获取正确的关系标签。

1 相关工作

在自然语言处理(NLP)的各项任务中,要把自然语言转化成机器能够理解的语言,就需要把单词向量化。目前的研究表明,不论是用于问答系统或是语义角色标注等任务,实现单词的向量化已经成为预处理部分的必经过程。2013年 Mikolov 等^[2]发布了 Word2Vec 工具,提供了 Skip-gram 和 CBOW 两种词向量训练模型,分别利用目标词预测上下文和利用上下文预测目标词。2016年 Facebook 开源了 FastText 工具,在使用负例采样的 Skip-gram 模型基础上,将每个中心词看作是子词的集合,学习子词的词向量^[3],学习速度快,效果不错。预训练的词向量^[4-5]能够从大规模的未标记文本中获取词汇的句法和语义信息,但是这些方法只允许每个词有一个上下文独立的与语境无关的表示。为了解决传统词向量的缺点,Neelakantan

等^[6]提出学习每个单词不同意义的不同向量。Wieting 等^[7]和 Bojanowski 等^[8]提出用子词信息来丰富词向量。除此之外,也有集中在学习上下文表示上的方法。2016年 Melamud 等^[9]使用 BiLSTM 从大型语料库中对一个目标词进行周围上下文的编码。其他计算上下文嵌入的方法包含目标词本身,编码器计算方式是远程神经机器翻译(COVE)^[10]或无监督语言模型^[11]。上述方法都依赖于大型语料库,且生成单一的向量表示。基于此,2018年 Matthew 等提出了以深层双向语言模型(BiLM)为基础,用各层之间的线性组合来表示词向量的方式,称为 ELMo 模型,解决单词在语义和语法上的复杂特点,可以根据上下文动态地生成词向量,解决一词多义问题^[12]。本文借鉴该模型来处理藏文语义歧义问题。

关系抽取是自然语言处理中最重要的任务之一。目前,大量的工作都是基于有监督的方法用于关系抽取,但是都需要人工大规模的标记数据,费时费力。近年来,深度学习广泛应用于各种领域^[13],也成功地应用于不同的 NLP 任务,如词性标注^[14]、情感分析^[15]、句法分析^[16]和机器翻译^[17]等,因而许多研究人员利用神经网络来自动学习特征用于关系抽取,但是在句子层面提取关系,缺乏足够的训练数据。为了解决这个问题,Mintz 等提出了将远程监督方法应用到关系抽取任务上,将文本对齐到构建的知识库,并使用该对齐来训练关系抽取器。各种大规模的知识库如 YAGO^[18]、Freebase^[19]、DBpedia^[20]、NELL^[21]等被广泛构建并应用于关系抽取。Zeng 等不用 NLP 工具预处理,利用 PCNN 自动提取句子级别的特征,并考虑实体位置的结构信息,其 P@avg 达到 67.6%^[22],但多实例学习模块在训练过程中只能选择一个有效的句子,不能充分利用句子信息。2016年,Lin 等用卷积神经网络嵌入句子的语义,然后在多实例上构建句子级别的注意力机制,动态减少有噪声实例的权值,加强正样本,更加有效地利用了句子的信息,其 P@avg 达到 72.2%^[23]。Jiang 等考虑了实体对的多关系问题,在构建句子向量时直接在每一维度取所有句子对应维度的最大值,这样就融合了所有句子的信息,其 P@avg 达到 72%^[24]。Feng 等基于 Memory Network 的思想来进行关系抽取,引入了基于关系的注意力机制来判断关系之间的相似度,其 P@avg 达到 79.7%^[25]。Ji 等^[26]在实体关系抽取过程中加入了实体描述信息,即每个实体相关的文字描述,这样加强了对实体

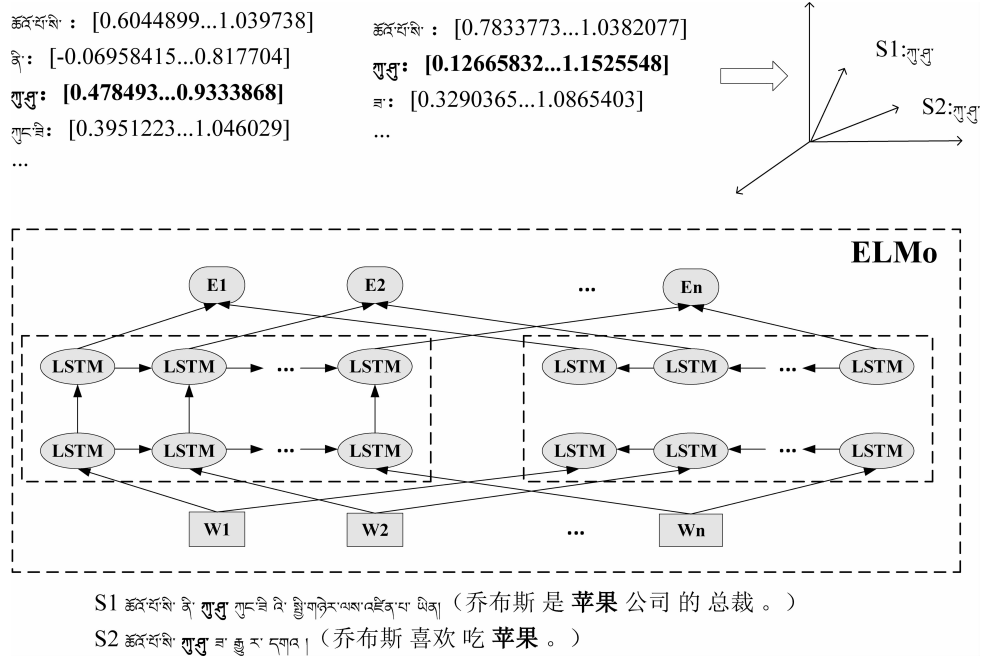


图 2 ELMo 结构示意图

$$p(t_1, t_2, \dots, t_N) = \prod_{k=1}^N p(t_k | t_1, t_2, \dots, t_{k-1}) \quad (1)$$

后向语言模型与之相反,是通过下文($t_{k+1}, t_{k+2}, \dots, t_N$)来预测之前第 k^{th} 词 t_k 的概率,如式(2)所示。

$$p(t_1, t_2, \dots, t_N) = \prod_{k=1}^N p(t_k | t_{k+1}, t_{k+2}, \dots, t_N) \quad (2)$$

ELMo 的基本框架是双向长短时记忆网络(BiLSTM),训练目标函数是将前后向语言模型结合起来,最大化前向、后向模型的联合似然函数即可,如式(3)所示。

$$\sum_{k=1}^N \left[\log p(t_k | t_1, t_2, \dots, t_{k-1}) + \log p(t_k | t_{k+1}, t_{k+2}, \dots, t_N) \right] \quad (3)$$

这样,句子中的每个单词都可以得到三个词向量 $\{e, \tilde{e}_{k,j}^{LM}, \tilde{e}_{k,j}^{LM} | j=1,2\}$ 。然后,将词向量视为下游任务的输入。为了应用到关系提取模型中,ELMo 通过计算不同层的任务特定权重将三个嵌入折叠成单个单词向量, $e_{k,j}^{LM} (j=0,1,2)$,如式(4)所示。

$$ELMo = \gamma \cdot \sum_{j=0}^2 \alpha_j \cdot e_{k,j}^{LM} \quad (4)$$

其中, α_j 是每层单词向量基于任务的权重, γ 是

用于调整关系提取任务中词向量权重的参数。

2.2 注意力机制

远程监控的方法是将纯文本中的实体与知识库中的实体对齐,将知识库中的实体对关系标记为语料库中的实体对关系。如果知识库中某些实体对之间没有关系,则将其标记为负实例(NA)。然而自动标记不可避免地会出现错误的标签。所以, Riedel 等^[30]提出采用多实例学习的方法来减少错误标记,将训练集 T 分解为多个实体对的包 $\{ \langle h_1, t_1 \rangle, \langle h_2, t_2 \rangle, \dots, \langle h_n, t_n \rangle \}$,每个包内的每个实体对包含 $\{s_1, s_2, \dots, s_m\}$ 个句子,这些句子都包含头实体 h_i 和尾实体 t_i 。Zeng 等^[22]尝试将多实例方法与神经网络模型相结合建立关系抽取器,但是在训练和测试时,对每个实体对只选择一个最有可能表示该包关系的句子,该方法会忽略未被选择的句子信息。

为了充分利用所有的句子信息,采用句子级别的注意力机制进行关系抽取。由于进行关系预测的关键信息可能出现在句子的任何位置,所以需要学习到句子的所有局部特征。给定一个句子 s 和两个目标实体 $\langle e_1, e_2 \rangle$,利用分段卷积神经网络(PCNN)构造句子的分布式表示 S ,其中每个句子的表示 x_i 包含实体对 $\langle e_1, e_2 \rangle$ 是否有关系 r 的信息,然后利用句子级别的注意力机制来判断每个句子的贡献度,集合向量 x 是包内所有句子的加权和,如

式(5)所示。

$$x = \sum_i \alpha_i \cdot S_i \quad (5)$$

其中, α_i 为每个句子表示 S_i 的权重, α_i 的计算如式(6)所示。

$$\alpha_i = \frac{\exp(e_i)}{\sum_k \exp(e_k)} \quad (6)$$

其中, e_i 是计算输入句子 s_i 与关系 r 的概率函数, e_i 的计算如式(7)所示。

$$e_i = S_i \cdot A \cdot r \quad (7)$$

其中, A 为加权对角矩阵, r 为关系 r 的向量表示。

2.3 联合得分函数

为了进一步减少错误标签问题,我们使用 soft-label 联合得分函数来修正实体对所在包的错误标签^[30]。Soft-label 方法是指在训练过程中利用相似的语义结构来纠正错误的远程监督(DS)标签,动态推导出每个包的标签,基于正确标记的实例,计算如式(8)所示。

$$r_i = \operatorname{argmax}(o + \max(o)A \odot L_i) \quad (8)$$

其中, L_i 表示 DS 标签, A 表示 DS 标签的可靠性, \odot 表示点积操作, o 表示基于 $\langle e_1, e_2 \rangle$ 实体对表示 s_i 的关系得分向量,计算式如(9)所示。

$$o_i = \frac{\exp(\mathbf{M} \cdot s_i + b)}{\sum_k \exp(\mathbf{M} \cdot s_k + b)} \quad (9)$$

其中, M 为关系矩阵, b 为偏置项。在训练过程中,利用实体对级别的交叉熵损失函数来计算得到的标签作为准标签的概率,计算如式(10)所示。

$$J(\theta) = \sum_{i=1}^n \log p(r_i | S_i; \theta) \quad (10)$$

3 实验

3.1 数据集

本文首先在 Mintz 等^[1]提出的英文数据集上进行试验。该数据集是使用 Freebase^[19]作为远程监督知识库和 New York Times (NYT) 语料库进行对齐。以 2005—2006 年 NYT 的句子作为训练集, 2007 年的句子作为测试集。其中训练数据包含 522 611 个句子、281 270 个实体对; 测试数据包含 172 448 个句子、96 678 个实体对, 选择 53 个关系来进行试验, 语料格式为六元组 $\langle e_{id1}, e_{id2}, e_1, e_2,$

$r, s \rangle$, 分别表示实体 1 的编号, 实体 2 的编号, 实体 1, 实体 2, 关系类型和包含该实体对的句子。

英文数据集的实验结果证明了方法的有效性, 接下来在藏文数据上进行研究。数据集是爬取的藏文网站语料, 构建了 103 509 条藏文知识库, 将爬取的语料与知识库对齐之后, 共挑选出 4 126 条有效句子, 其中 3 126 条用于训练, 1 000 条用于测试, 选择 11 个关系来进行试验, 数据格式同 NYT+Freebase 数据集, UUID 是每个实体的独一无二编号。

3.2 评测指标

本文采用传统的 F_1 值作为评价指标, 同时与前面的工作类似, 加入 top- N 精度 ($P@N$) 来进一步分析, 此评测方法是根据预测关系概率进行排序, N 为选择出来的句子数量。

3.3 实验设置与结果分析

本文进行了英文语料和藏文语料的对比实验, 在实验中, 设置了分段卷积神经网络(PCNN)和 ELMo 语言模型的参数, 如表 1 和表 2 所示。

表 1 PCNN 主要参数设置

Parameter	Value
Batch size	160
Window size	3
Word dimension	50
Filter dimension	230
Learning_rate	0.01

表 2 ELMo 主要参数设置

Parameter	Value
Hidden size	128
Batch size	200
Epoch	10
Learning_rate	0.001
Dropout	0.5

本文以分段卷积神经网络为基线, 用 F_1 值和 $P@N$ 参数来评估, 为了证明模型的有效性, 我们分别在英文和藏文语料上对于不同的模型均进行了对比实验, 实验结果如表 3 和表 4 所示。

表 3 词向量与语言模型在英文上的结果

模型		$F_1/\%$	$P@N/\%$			
			100	200	300	avg
Word2Vec+	PCNN	35.6	68.3	60.7	53.8	60.9
	+ selective-attention	44.5	72.3	69.7	64.1	68.7
	+ soft_label	47.9	76.2	73.1	67.4	72.2
ELMo+	PCNN	49.8	70.3	62.7	55.8	62.9
	+ selective-attention	53.5	77.0	72.5	67.7	72.4
	+ soft_label	61.2	80.0	74.5	69.7	74.7

表 4 词向量与语言模型在藏文上的结果

模型		$F_1/\%$	$P@N/\%$			
			100	200	300	avg
Word2Vec+	PCNN	34.7	48.6	44.0	41.8	44.8
	+ selective-attention	42.9	51.9	49.1	47.4	49.5
	+ soft_label	47.2	56.3	51.2	50.8	52.7
ELMo+	PCNN	41.2	60.9	55.8	53.8	56.8
	+ selective-attention	49.3	64.4	60.2	60.1	60.4
	+ soft_label	58.9	70.2	67.2	58.8	67.1

由表 3 可以看出,在英文语料规模上,基于语言模型的方法在 F_1 值达到 61.2%, $P@N$ 达到 74.7%,都优于 Word2Vec 词向量的效果,说明语言模型确实可以解决语义歧义的问题。在模型效果上可以看出,我们提出的方法比基线模型效果提升了 25.6%,说明我们的方法可以有效地提高关系抽取的准确率。具体分析如下:在传统的词向量模型和 ELMo 语言模型中,加入选择性注意力机制之后,效果分别有了 8.9%和 3.7%的提升,说明注意力机制可以提高实验效果,更好地学习句子的特征。语言模型的提升比 Word2Vec 的小,可能原因是语言模型已经初步提取过句子的语法和语义信息,在语义方面已经筛选过。在引入 soft-label 方法后,在传统的词向量模型和 ELMo 语言模型中,效果分别有了 3.4%和 7.7%的提升,证明了该方法的有效性,缓解了错误标记问题。综上所述,我们提出的方法在实体关系抽取任务是可行的,可以用在藏文的实体关系抽取任务中。

由表 4 可以看出,在藏文语料中的实验结果总体趋势与英文语料的结果一致。基于语言模型的方法在 F_1 值达到 58.9%, $P@N$ 达到 67.1%,比传统的 Word2Vec 词向量模型效果分别提升了 24.2%

和 22.3%,说明我们提出的改进方法在减少人工标注语料的情况下,可以提高藏文实体关系抽取的准确率。具体分析如下:在传统的词向量模型和 ELMo 语言模型中,加入选择性注意力机制之后,效果分别有了 8.1%和 7.1%的提升,说明注意力机制可以减小噪声实例的权重,学到正例句子的特征。引入 soft-label 方法后,在传统的词向量模型和 ELMo 语言模型中,效果分别有了 4.3%和 9.6%的提升,证明 soft-label 方法在实体对级别缓解了错误标记问题。综上所述,我们提出的方法可以提高藏文实体关系抽取的准确率。

藏文实体关系抽取方法比较如表 5 所示。由表 5 可以看出,我们提出的模型的准确率比联合模型的效果高 2.9%,主要原因是在预处理阶段使用 ELMo 生成动态词向量,降低了语义歧义的影响;选择性注意力机制可有效提取句子特征;soft-label 标签方法减轻了错误的标签问题。这说明改进的远程监督是一种有效的方法。但是与传统的基于人工标注语料的方法相比,我们的实验结果略逊色,主要原因是在实验过程中未对藏文语料进行处理,单纯的与知识库匹配之后的语料作为训练语料,其中包含大部分只有单个句子的包,这会影响关系抽取的效

果。其次,本次实验未考虑藏文的语言特征,未来会针对藏文语料特点,加入对语料的处理。

表 5 藏文实体关系抽取方法比较

模型	$F_1/\%$
SVM ^[27]	62.6
BP ^[28]	62.1
联合模型 ^[29]	56.0
Ours	58.9

综合表 3 和表 4,我们发现藏文的整体效果没有英文好,经过分析,原因如下:①英文知识库构建比较完善,包含 53 种关系类型,而藏文知识库规模较小,仅有 11 种关系类型;②英文的语料容易获取,语料量大且质量较高,可以得到客观的实验结果,而藏文知识库规模较小,语言处理也不方便;③在实验中,未考虑藏文的语言特征,忽略了语言特征对实验效果的作用。

4 总结与展望

本文主要介绍了针对藏文的实体关系抽取方法,提出在预处理阶段加入语言模型来缓解一词多义的问题,在训练过程中加入选择性注意力机制,来获取包内句子的权重以及对包的贡献度,之后加入联合得分函数,动态修正远程监督的错误标签问题,最终的实验结果 F_1 值为 58.9%, $P@N$ 为 67.1%,较之前的方法均有一定提升。

但本文的方法没有考虑藏文的语法特征,在语料规模上还存在不足。在未来的工作中,我们会加大藏文的语料规模,加入藏文的语法特征,优化对藏文的处理,同时还将结合强化学习方法在句子级别减少错误标记问题,进行去噪,以提高藏文实体关系提取的准确性。

参考文献

[1] M. Mintz, S. Bills, R. Snow, et al. Distant supervision for relation extraction without labeled data[C]// Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009: 1003-1011.

[2] T. Mikolov, K. Chen, G. Corrado, et al. Efficient estimation of word representations in vector space[J].

arXiv preprint arXiv: 1301.3781, 2013.

[3] E Grave, T Mikolov, A Joulin, et al. Bag of tricks for efficient text classification[J]. arXiv preprint arXiv: 1607.01759, 2016.

[4] J P Turian, L A Ratinov, Y Bengio. Word representations: A simple and general method for semi-supervised learning[C]// Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010: 384-394.

[5] T Mikolov, I Sutskever, K Chen, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26: 3111-3119.

[6] A Neelakantan, J Shankar, A Passos, et al. Efficient non-parametric estimation of multiple embeddings per word in vector space[C]// Proceedings of the Empirical Methods in Natural Language Processing, 2015: 1059-1069.

[7] J Wieting, M Bansal, K Gimpel, et al. Charagram: Embedding words and sentences via character n-grams [C]// Proceedings of the Empirical Methods in Natural Language Processing, 2016: 1504-1515.

[8] P Bojanowski, E Grave, A Joulin, et al. Enriching word vectors with subword information[J]. Transactions of the Association for Computational Linguistics, 2017, 5: 135-146.

[9] O Melamud, J Goldberger, I Dagan. Context2vec: Learning generic context embedding with bidirectional LSTM[C]// Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, 2016: 51-61.

[10] B Mccann, J Bradbury, C M. Xiong, et al. Learned in translation: Contextualized word vectors [C]// Proceedings of the Neural Information Processing Systems, 2017: 6294-6305.

[11] M E Peters, W Ammar, C Bhagavatula, et al. Semi-supervised sequence tagging with bidirectional language models[C]// Proceedings of the Association for Computational Linguistics, 2017: 1756-1765.

[12] M E Peters, M Neumann, M Iyyer, et al. Deep contextualized word representations[C]// Proceedings of the Association for Computational Linguistics, 2018: 2227-2237.

[13] Y Bengio. Learning Deep architectures for AI[J]. Foundations and Trends R in Machine Learning, 2009, 2(1): 1-127.

[14] R Collobert, J Weston, L Bottou, et al. Natural language processing (almost) from scratch[J]. Journal of Machine Learning Research, 2011, 12(1): 2493-2537.

[15] C N dos Santos, M A de C Gatti. Deep convolutional

- neural networks for sentiment analysis of short texts [C]//Proceedings of the COLING 2014, the 25th International Conference on Computational Linguistics, 2014: 69-78.
- [16] R Socher, J Bauer, C D Manning, et al. Parsing with compositional vector grammars[C]// Proceedings of the Association for Computational Linguistics. 2013: 455-465.
- [17] I Sutskever, O Vinyals, Q V Le. Sequence to sequence learning with neural networks[J]. Advances in Neural Information Processing Systems, 2014, 2: 3104-3112.
- [18] J Hoffart, F M Suchanek, K Berberich, et al. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia [J]. Artificial Intelligence, 2013, 194: 28-61.
- [19] B Yue, M Gui, J H. Guo, et al. An effective framework for question answering over Freebase via reconstructing natural sequences[C]// Proceedings of The 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, 2017: 865-866.
- [20] D Ritze, C Bizer. Matching web tables to DBpedia—a feature utility study[C]// Proceedings of the 20th EDBT, 2017: 210-221.
- [21] F A O Santos, F B do Nascimento, M I S Santos, et al. Training neural tensor networks with the never ending language learner[M]. Information Technology-New Generations. Cham: Springer, 2018: 19-23.
- [22] D J Zeng, K Liu, Y B Chen, et al. Distant supervision for relation extraction via piecewise convolutional, neural networks[C]// Proceedings of the Empirical Methods in Natural Language Processing, 2015: 1753-1762.
- [23] Y K Lin, S Q Shen, Z Y Liu, et al. Neural relation extraction with selective attention over instances [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 2124-2133.
- [24] X Jiang, Q Wang, P Li, et al. Relation extraction with multi-instance multi-label convolutional neural networks[C] //Proceedings of the COLING, 2016: 1471-1480.
- [25] X C Feng, J Guo, B Qin, et al. Effective deep memory networks for distant supervised relation extraction [C] //Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. 2017: 4002-4008.
- [26] G L Ji, K Liu, S Z He, et al. Distant supervision for relation extraction with sentence-level attention and entity descriptions[C] // Proceedings of the AAAI, 2017: 3060-3066.
- [27] 朱臻, 孙媛. 基于 SVM 和泛化模板协作的藏语人物属性抽取[J]. 中文信息学报, 2015, 29(6): 220-227.
- [28] 郭莉莉, 孙媛. 基于 BP 神经网络的藏语实体关系抽取[J/OL]. 软件导刊, 2019, 18(03): 13-15, 21.
- [29] 夏天赐, 孙媛. 基于联合模型的藏文实体关系抽取方法研究[J]. 中文信息学报, 2018, 32(12): 76-83.
- [30] S Riedel, L M Yao, A McCallum. Modeling relations and their mentions without labeled text[C]// Proceedings of The ECML PKDD, 2010: 148-163.
- [31] T Y Liu, K X Wang, B B Chang, 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.



王丽客(1996—), 硕士研究生, 主要研究领域为自然语言处理和知识图谱。
E-mail: 18366184191@163.com



夏天赐(1993—), 硕士研究生, 主要研究领域为自然语言处理、信息检索和问答系统。
E-mail: muctianciking@163.com



孙媛(1979—), 通信作者, 博士, 副教授, 主要研究领域为自然语言处理和知识图谱。
E-mail: tracy.yuan.sun@gmail.com