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。远程监督用于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，因而可以用在缺少训练语料的藏文领域。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。本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的准确率，且优于基线模型效果。

关键字: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: The task of relation extraction is classifying the relation between two entities in a sentence. Distant supervision for relation extraction is an efficient method to automatically aligns entities in texts to a given KB, which alleviated the problem of manual labelling. However, there are wrong label problems, a lot of noise occurs when extracting features and it is unable to learn valid information. In this paper, we propose an improved distant supervised relation extraction model in Tibetan based on Piecewise Convolutional Neural Network (PCNN) to alleviate wrong labelling problems and extract effective features automatically which combines language model with selective-attention mechanism. Soft-label method is introduced to dynamically correct the relation label. The experimental results show that our method is effective and outperforms several competitive baseline methods.

Key words: 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的标注数据集，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 等人提出远程监督的方法[1]，通过对齐知识库与文本，来自动生成大量的训练数据，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实体关系抽取任务上，即对于知识库中的三元组$<e_1,r,e_2>$，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。远程监督刚好弥补这一缺陷，可以自动标注训练语料。如知识库中有实体对<ཚའུ་བསི་（乔布斯），གུ་ཤུ（苹果）>，在语料库中包含该实体对的句子可能有多个，如：**ཚའུ་བསི་** ནི་ **གུ་ཤུ** གུང་ཟེ་ཤེ་ སྦྱི་གཉེར་ལས་ཁཐོན་པ་ ཡིན། （**乔布斯**是**苹果**公司的总裁。）和 **ཚའུ་བསི་ གུ་ཤུ** ར་ ཐུ་ རང་དགའ། （**乔布斯**喜欢吃**苹果**）。远程监督方法会将所有包含该实体对<ཚའུ་བསི་（乔布斯），གུ་ཤུ（苹果）>的句子都提取出来作为训练语句。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“གུ་ཤུ （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，为了更有效的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 等发布了 word2vec 工具[2]，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表示。为了解决传统词向量的缺点，Neelakantan 等人提出学习每个单词不同意义的不同向量[6]. Wieting 等人[7]和 Bojanowski 等人[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 等人使用 BiLSTM 从大型语料库中对一个目标词进行周围上下文的编码[9]。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 等人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的生成词向量，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，但是都需要大量的标记数据，费时费力。近年来，深度学习广泛应用于各种领域[13]，也成功的应用于不同的 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 等人提出了将远程监督方法应用到关系抽取上，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 等人不用 NLP 工具预处理，利用 PCNN 自动提取句子级别的特征，考虑实体位置的结构信息，P@avg 达到 67.6%[22]。但其多实

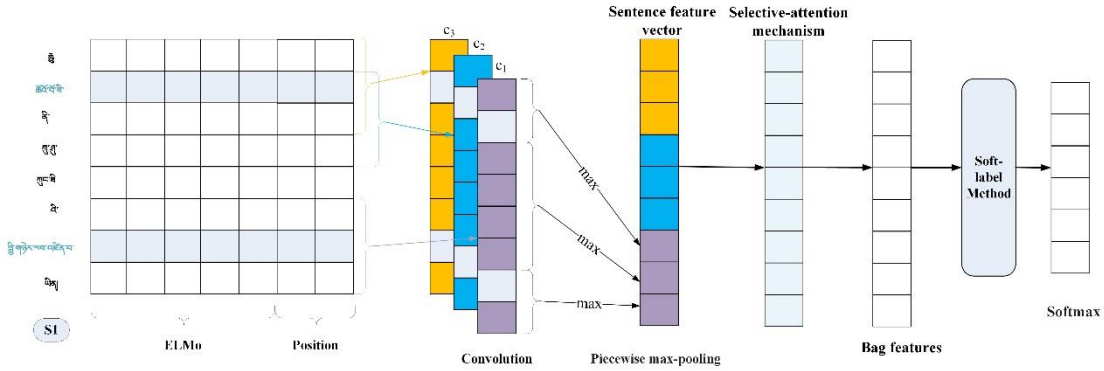

图 1 整体框架图

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 等人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，更有效的利用了句子的信息，P@avg 达到 72.2%[23]。Jiang 等人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 sentence 对应维度的最大值，这样就融合了所有句子的信息，P@avg 达到 72%[24]。Feng 等人基于 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，P@avg 达到 79.7%[25]。Ji 等人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，这样加强了对实体向量的学习，P@avg 达到 81.3%[26]。

如上描述均是针对英文领域开放数据集（NYT + FreeBase）进行的实体关系抽取研究，对本文的研究内容而言，接下来介绍藏文领域实体关系抽取的研究现状。在藏文关系抽取方面的资料可查询的较少，主要有朱臻等提出基于泛化模板与 SVM 相结合的方法抽取藏文人物属性，在属性类别中 F1 值最高达到 62.61%[27]。郭莉莉等基于 BP 神经网络，在预处理阶段添加实体关系，实体距离关系，实体及周围词特征，在 4216 句实验语料中 F1 值达到 62.12%[28]。夏天赐等基于联合模型方法，在字和词级别对藏语进行预处理，加入词性标注特征，将实体关系抽取任务转变为序列标注的问题，在 2400 句实验语料中 F1 值达到 56%[29]。

以上关于藏文关系抽取的研究都是基于人工标注的语料，且语料规模较小。基于以上研究内容，同时考虑藏文语料规模较小，以及存在歧义等问题，为了更好的学习句子的信息，进而解决错误标注问题，本文首先加入语言模型生成动态词向量，然后加入选择性注意力机制计算句子对包内关系的贡献度，最后加入联合评分函数，在训练过程中动态获取关系标签，来提高藏文实体关系抽取的准确率。

2. 方法与模型

为了自动标注训练语料库，缓解错误标注问题，本文提出了基于分段卷积神经网络的远程监督藏文实体关系抽取方法的改进模型。在对语料库进行预处理之后，我们使用 ELMo 生成动态单词向量，然后将单词向量和位置向量发送到神经网络层以生成句子向量。在卷积层中，我们设置滑动窗口的大小来自动提取句子的局部特征。此处，我们加入选择性注意机制来获得一个包中每个句子的权重，然后提取包的特征。最后，引入 soft-label 去噪函数来动态修正实体对级别的错误标签，然后使用 softmax 分类器计算每个关系的置信度。整体框架如图 1 所示，下面具体介绍每一部分的内容。

2.1 语言模型

在获取训练语料之后，首先对语料进行分句，分词等预处理，在与知识库对齐过程中，已经识别出句中的实体，其中至少出现两个实体的句子作为有效语料；此时需要训练词向量作为卷积层的输入。在藏文语料中，存在

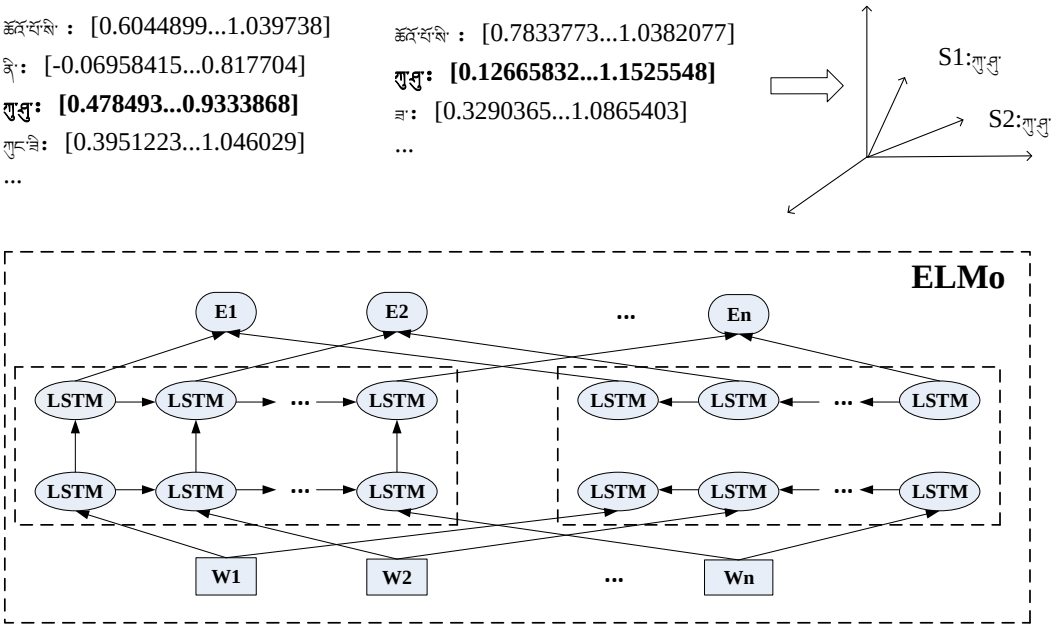

图 2 ELMo 结构示意图

语义不明确，一词多义的问题，而传统的词向量训练模型在对多义单词进行编码时，无法区分含义，不同的上下文信息会编码到相同的词向量空间，因而无法区分多义词的同语义只是对每个词生成了一个静态的词向量，造成词向量的质量相对较差。针对这个问题，我们借鉴了 Matthew 等人提出的双向语言模型 ELMo，其本质思想是：事先用语言模型学好单词的向量表示，此时多义词无法区分，但是在实际使用词向量的时候，单词已经具备特定的上下文，可以根据上下文单词的语义去调整单词的向量表示，经过调整更能表达在上下文中的具体含义，自然也就解决了多义词的问题。ELMo 在中文和英文上针对不同任务均有不同程度的提升，所以可以用来解决藏文中存在的一词多义问题。

图 2 是 ELMo 的结构图。ELMo 包含一个两层双向语言模型和一层字符卷积层。双向语言模型包括前向语言模型和后向语言模型，前向语言模型就是已知$(t_1,t_2,...,t_{k-1})$，预测之后t_k的概率，如公式（1）所示。

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

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

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

ELMo 的基本框架是双层长短时记忆网络（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，如公式（3）所示。

$$\sum_{k=1}^{N}\left(\begin{array}{l}\log p(t_k|t_1,t_2,...,t_{k-1})+\\ \log p(t_k|t_{k+1},t_{k+2},...,t_N)\end{array}\right) \quad (3)$$

这样，句子中的每个单词都可以得到三个词向量$\left\{e,\overrightarrow{e}_{k,j}^{LM},\overleftarrow{e}_{k,j}^{LM}|j=1,2\right\}$。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，$e_{k,j}^{LM}(j=0,1,2)$，如公式（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会出现错误的标签。所以，Riedel 等人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$\{<h_1,t_1>,<h_2,t_2>,...,<h_n,t_n>\}$，每个包内的每个实体对包含$\{s_1,s_2,...,s_m\}$个句子，这些句子都包含头实体$h_i$和尾实体$t_i$[30]。Zeng 等人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[22]，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$<e_1,e_2>$，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$<e_1,e_2>$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，如公式（5）所示。

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

其中α_i为每个句子表示S_i的权重，α_i的计算如公式（3）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的计算如公式（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，计算公式如（8）所示。

$$r_i = arg\max(o + \max(o) A \odot L_i) \quad (8)$$

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$<e_1,e_2>$实体对表示s_i的关系得分向量，计算公式如（9）所示。

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

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如公式（10）所示来计算得到的标签作为准标签的概率。

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

3. 实验

3.1 数据集

本文首先在 Mintz 等人提出的基准英文数据集上进行试验[1]。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 522611 个句子、281270 个实体对；测试数据包含 172448 个句子、96678 个实体对，选择 53 个关系来进行试验，语料格式为六元组$<e_{id1},e_{id2},e_1,e_2,r,s>$，分别表示实体 1 的

编号，实体 2 的编号，实体 1，实体 2，关系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 F1 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 2 和 3 所示，语言模型生成的应用到下游任务的文件如表 4 所示。

表 2 PCNN 主要参数设置

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

表 3 ELMo 主要参数设置

Paramter	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 F1 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 4 和 5 所示：

由表 4 可以看出，在英文语料规模上，基于语言模型的方法在 F1 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的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，更好的学习句子的特征。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上是可行的，可以用在藏文的实体关系抽取任务中。

由表 5 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 F1 值达到 58.9%，P@avg 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。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 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。

表 4 词向量与语言模型在英文上的结果

模型		F1（%）	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 5 词向量与语言模型在藏文上的结果

模型		F1（%）	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**

表 6 藏文实体关系抽取方法比较

方法	F1（%）
SVM[27]	0.626
BP[28]	0.621
联合模型[29]	0.560
Ours	**0.589**

由表 6 可以看出，我们提出的模型的准确率比联合模型的效果搞 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，本次实验未考虑到藏文的语言特征，未来会针对藏文语料特点，加入对语料的处理。

综合表 4 和表 5，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，加入语言模型来减少一词多义的问题，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 F1 值为 58.9%，P@avg 为 67.1%，较之前的方法有了一些提升。

但是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，同时我们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] Mintz, M, Bills, S, Snow, R, 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] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[J]. Computer Science, 2013.

[3] Joulin A, Grave E, Bojanowski P, et al. Bag of Tricks for Efficient Text Classification[J]. Proceedings of arXiv 2016.

[4] Turian J P, Ratinov L A, Bengio Y. 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.

[5] Mikolov T, Sutskever I, Chen K, et al. Distributed Representations of Words and Phrases and their Compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26:3111-3119.

[6] Neelakantan A, Shankar J, Passos A, et al. Efficient Non-parametric Estimation of Multiple Embeddings per Word in Vector Space[J]. Computer Science, 2015.

[7] Wieting J, Bansal M, Gimpel K, et al. Charagram: Embedding Words and Sentences via Character n-grams[J]. 2016.

[8] Bojanowski P, Grave E, Joulin A, et al. Enriching Word Vectors with Subword Information[J]. Transactions of the Association for Computational Linguistics, 2017, 5:135-146.

[9] Melamud O, Goldberger J, Dagan I. context2vec: Learning Generic Context Embedding with (41): 19.

[21] Santos F A O, do Nascimento F B, Santos M S, et al. Training neural tensor networks with the never ending language learner [J]// Information Technology-Bidirectional LSTM[C]//Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning. 2016.

[10] Mccann B , Bradbury J , Xiong C , et al. Learned in Translation: Contextualized Word Vectors[J]. In NIPS 2017.

[11] Matthew E, Peters, Waleed Ammar, et al. Semi-supervised sequence tagging with bidirectional language models[C]// Proceedings of The Association for Computational Linguistics. 2017.

[12] Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations[J]. 2018.

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

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

[15] dos Santos, C. N. & Gatti, M. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. 2014.

[16] Socher R, Bauer J, Manning C D, et al. Parsing With Compositional Vector Grammars[C]// Meeting of the Association for Computational Linguistics. 2013.

[17] Sutskever I, Vinyals O, Le Q V. Sequence to Sequence Learning with Neural Networks[J]. 2014.

[18] Hoffart J, Suchanek F M, Berberich K, et al. YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia [J]. Artificial Intelligence, 2013, 194: 28-61.

[19] Yue B, Gui M, Guo J, et al. An Effective Framework for Question Answering over Freebase via Reconstructing Natural Sequences[C]// Proceedings of the 26th International Conference. International World Wide Web Conferences Steering Committee, 2017: 865-866.

[20] Ritze D, Bizer C. Matching Web tables to DBpedia-a feature utility study [J]. Context, 2017, 42

New Generations. Cham: Springer, 2018: 19-23.

[22] Zeng D , Liu K , Chen Y , et al. Distant Supervision for Relation Extraction via Piecewise Convolutional, Neural Networks[C]// Conference on Empirical

Methods in Natural Language Processing. 2015.

[23] Yankai Lin, Shiqi Shen, Zhiyuan Liu, et al. Neural relation extraction with selective attention over instances[C]// Proceedings of The Association for Computational Linguistics.2016.

[24] X. Jiang, Q. Wang, P. Li, and B. Wang. Relation Extraction with Multi-instance Multi-label Convolutional Neural Networks[C]. In COLING, 2016.

[25] X. 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] Guoliang Ji, Kang Liu, Shizhu He, et al. Distant supervision for relation extraction with sentence-level attention and entity descriptions[C] In 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).

[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, 2018, 1-4.

[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).

[30] Riedel S , Yao L , Mccallum A . Modeling Relations and Their Mentions without Labeled Text[C]// Machine Learning and Knowledge Discovery in Databases, European Conference, 2010, 20-24.

[31] Liu T , Wang K , Chang B , et al. A Soft-label Method for Noise-tolerant Distantly Supervised Relation Extraction[C]// Conference on Empirical Methods in Natural Language Processing. 2017.

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

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

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