

结合残差 BiLSTM 与句袋注意力的远程监督关系抽取

江 旭, 钱雪忠, 宋 威

(江南大学 人工智能与计算机学院, 江苏 无锡 214112)

摘 要: 实体关系抽取是从非结构化和程序化的海量文本中识别出实体之间的语义关系, 为本体构建、智能检索等任务提供数据支持, 然而现有远程监督关系抽取方法普遍存在需要大量人工标注语料库、提取特征含有噪声且忽略了实体与句子之间关联关系等问题。提出一种基于残差双向长短时记忆网络(BiLSTM)与句袋内和句袋间注意力机制的关系抽取模型, 在将词向量和位置向量作为模型输入的基础上, 通过残差 BiLSTM 网络提取语句与实体词中的长距离文本信息, 利用句袋内和句袋间注意力机制对提取到的特征信息进行处理, 使模型在远程监督过程中减少实体之间的特征提取噪声, 并提高模型识别准确性。在 NYT 数据集上的实验结果表明, 该模型能够充分利用实体与关系特征, 平均精确率达到 86.2%, 相比于将卷积神经网络和分段卷积神经网络作为句子编码器的同类模型具有更好的远程监督关系抽取性能。

关键词: 自然语言处理; 关系抽取; 残差连接; 双向长短时记忆网络; 句袋注意力机制

开放科学(资源服务)标志码(OSID):



中文引用格式: 江旭, 钱雪忠, 宋威. 结合残差 BiLSTM 与句袋注意力的远程监督关系抽取[J]. 计算机工程, 2022, 48(10): 110-115, 122.

英文引用格式: JIANG X, QIAN X Z, SONG W. Distantly supervised relationship extraction combined with residual BiLSTM and sentence bag attention[J]. Computer Engineering, 2022, 48(10): 110-115, 122.

Distantly Supervised Relationship Extraction Combined with Residual BiLSTM and Sentence Bag Attention

JIANG Xu, QIAN Xuezhong, SONG Wei

(School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu 214122, China)

[Abstract] Entity relationship extraction entails identifying the semantic relationship between entities from unstructured and programmed massive texts and can provide data support for ontology construction, intelligent retrieval, and other tasks. However, the existing distantly supervised relationship extraction methods generally suffer from three problems: the need for numerous manual annotation corpora, noise contained in the extraction features, and the relationship between entities and sentences being ignored. This paper proposes a relationship extraction model based on a residual Bi-directional Long Short-Term Memory (BiLSTM) network and intra-sentence and inter-sentence bag attention mechanisms. Based on using word vectors and position vectors as model input, the model extracts long-distance text information in sentences and entity words through a residual BiLSTM network, and it uses the intra-sentence and inter-sentence bag attention mechanisms to process the extracted feature information. This enables the model to reduce the feature extraction noise between entities in the distantly supervision process and improve its recognition accuracy. The experimental results on the New York Times (NYT) dataset demonstrate that the model can make full use of the entity and relationship features, and the average precision reaches 86.2%. Compared with similar models using Convolutional Neural Network (CNN) and Piecewise Convolutional Neural Network (PCNN) as sentence encoders, the proposed model has better remote supervised relationship extraction performance.

[Key words] natural language processing; relation extraction; residual connection; Bi-directional Long Short-Term Memory (BiLSTM) network; sentence bag attention mechanism

DOI: 10.19678/j.issn.1000-3428.0062891

0 概述

实体关系抽取作为信息抽取、自然语言理解、信息检索等领域的核心任务和重要环节, 可从非结构化和程序化的文本中提取实体之间的语义关系, 为用户提

供更加精准全面的信息。实体关系抽取通常是在非结构化和程序化的文本中提取二元关系并组成关系三元组的形式, 例如<Entity1, Relation, Entity2>, 其中, Entity1 和 Entity2 表示两个实体, Relation 表示两个实体之间的关系。给定一个句子“<e1>雷军</e1>创立了<e2>小米

基金项目: 国家自然科学基金(62076110); 江苏省自然科学基金(BK20181341)。

作者简介: 江 旭(1995—), 男, 硕士研究生, 主研方向为深度学习、数据挖掘; 钱雪忠, 副教授、硕士; 宋 威, 教授、博士。

收稿日期: 2021-10-08 **修回日期:** 2021-12-13 **E-mail:** 15036432197@163.com

科技有限责任公司</e2>”,可以看出两个实体之间的关系为“创立”。

在现有关系抽取方法中,监督关系抽取的准确率较高,但耗费大量人力资源。远程监督关系抽取方法通过数据自动对齐方式解决了大量无标签数据的自动标注问题,并且能够大幅减少标注成本,但存在知识库标注的句子有噪声、实体与关系之间表示不明确、无法准确表达句子与实体之间关系等问题。针对这些问题,研究人员提出了一系列解决方案。文献[1]提出图卷积网络,能够处理具有广义拓扑图结构的数据,并深入挖掘其实体和关系特征。文献[2]结合多实例与分段卷积神经网络(Piecewise Convolutional Neural Network, PCNN)进行远程监督关系抽取。文献[3]引入注意力机制,利用句子与关系来分配权重,通过对正确的句子与关系分配较高的权重,提升了关系抽取性能。随着深度学习技术的发展,神经网络^[4-5]被广泛应用于远程监督关系抽取。文献[6]提出分段卷积神经网络来建模句子表示,并选择准确的句子作为句袋表示。文献[7]使用多种神经网络作为句子编码器,并提出一种句袋内注意力机制,通过句袋中所有句子表示的加权和来计算句袋。文献[8]采用类似的注意力机制,并结合实体描述来计算权重。文献[9]提出一种软标签方法来降低噪声实例的影响。文献[10]采用双向长短期记忆(Bi-directional Long Short-Term Memory, BiLSTM)网络来提取句子特征,使用注意力机制来识别噪声句子以及句袋。

上述远程监督关系抽取方法利用句子嵌入的加权和来表示句袋^[11],以关系感知的方式计算句袋内的注意力权重,并在训练阶段使用相同的句袋表示来计算该句袋被分类到每个关系中的概率,然而这

种采取先识别实体再预测关系的端到端抽取方式会导致前一个任务的错误传递到下一个任务,并且忽略了实体与各个关系之间的联系。为解决上述问题,本文提出基于残差BiLSTM(ResNet_BiLSTM)与句袋内和句袋间注意力机制的实体关系抽取模型。通过句子嵌入的加权和计算关系感知句袋,并结合句袋注意力模块,在模型训练过程中动态计算句袋注意力权重,以解决句袋噪声问题。

1 基于ResNet_BiLSTM与句袋注意力的关系抽取

本文提出一种基于ResNet_BiLSTM与句袋注意力的关系抽取模型,用于远程监督关系抽取。 $\mathbf{g} = \{\mathbf{b}^1, \mathbf{b}^2, \dots, \mathbf{b}^n\}$ 表示一组远程监督给出的具有相同关系标签的句袋, n 是句袋数量, $\mathbf{b}^i = \{\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_m^i\}$ 表示句子数为 m 的句袋, $\mathbf{x}_j^i = \{\mathbf{w}_{j1}^i, \mathbf{w}_{j2}^i, \dots, \mathbf{w}_{jm}^i\}$ 表示在第 i 个句袋中的第 j 个句子, jm 表示第 j 个句子长度,模型框架如图1所示,主要包括以下模块:

- 1) 句子编码器。给定一个句子和句子中两个实体的位置^[12],得到句子的输入表示。
- 2) ResNet_BiLSTM特征提取器。由句子编码器得到的输入句子表示,通过输入ResNet_BiLSTM得到句子特征^[13]表示。
- 3) 句袋内注意力机制。给定句袋 \mathbf{b}^i 中所有句子的句子表示和关系嵌入矩阵 \mathbf{R} 、注意力机制权重向量 α_k^i 和句袋表示 \mathbf{b}_k^i 来计算所有关系,其中 k 为关系索引。
- 4) 句袋间注意力机制。给定一组句袋 \mathbf{g} ,通过基于相似性的注意力机制来进一步计算权重矩阵 β ,得到句袋组的表示。

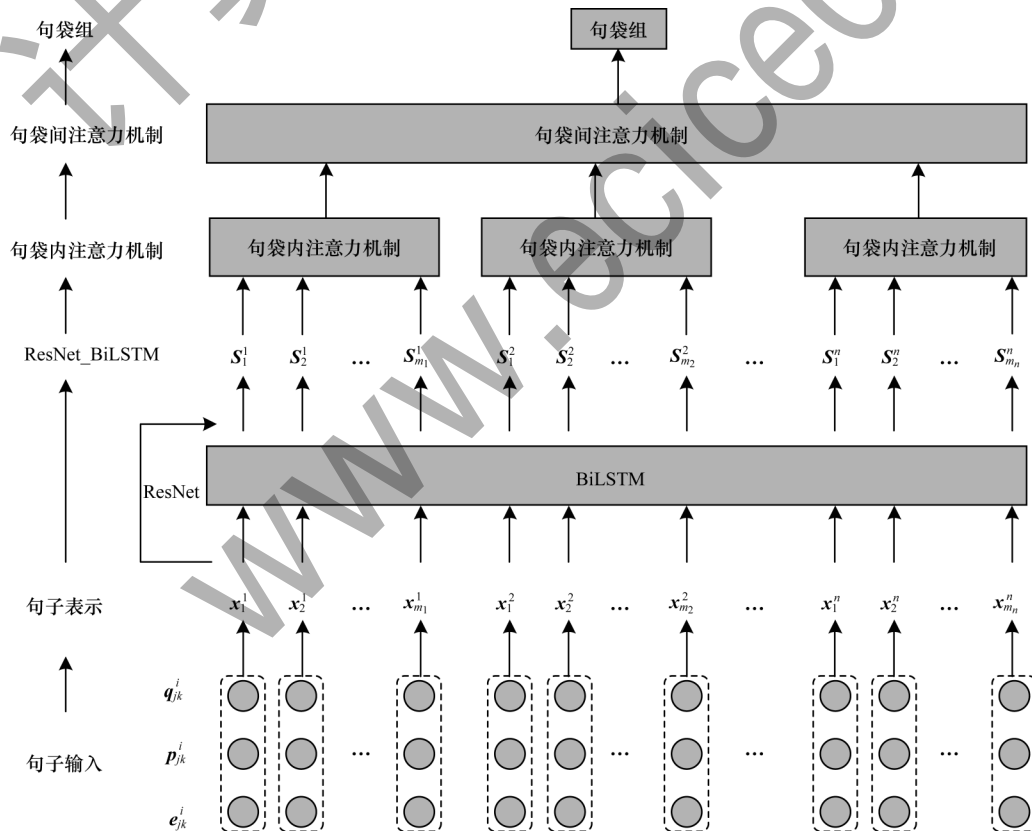


图1 基于ResNet_BiLSTM与句袋注意力的关系抽取模型框架

Fig.1 Framework of relationship extraction model based on ResNet_BiLSTM and sentence bag attention mechanism

1.1 句子编码器

句子的特征编码由词和词的位置特征表示,在句子 x_j^i 中每个词 w_{jk}^i 被映射成一个 d_w 维度的词嵌入,句子的特征向量表示为 e_{jk}^i ,位置特征^[14]是每个词到实体之间的距离,表示为 p_{jk}^i 和 q_{jk}^i ,将位置特征映射成 d_p 维的词嵌入,这3个向量的连接向量为 $d_w + 2d_p$ 维的向量,表示为 $w_{jk}^i = [e_{jk}^i, p_{jk}^i, q_{jk}^i]$ 。

1.2 ResNet_BiLSTM 特征提取器

首先,输入句子 x_j^i ,词的嵌入矩阵 $w_j^i \in R^{l_y \times (d_w + 2d_p)}$ 输入 ResNet_BiLSTM,如式(1)~式(5)所示:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (2)$$

$$g_t = \tanh(W_{xe}x_t + W_{he}h_{t-1} + W_{ce}c_{t-1} + b_e) \quad (3)$$

$$c_t = i_t g_t + f_t c_{t-1} \quad (4)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (5)$$

$$h_t = o_t \tanh(c_t) \quad (6)$$

其中:输入门 i_t 的权重为 $W_{xi}, W_{hi}, W_{ci}, b_i$;遗忘门 f_t 的权重为 $W_{xf}, W_{hf}, W_{cf}, b_f$;输出门 o_t 的权重为 $W_{xo}, W_{ho}, W_{co}, b_o$ 。

然后,ResNet_BiLSTM 计算如下:

$$H(x) = F(x) + x \quad (7)$$

其中: $F(x)$ 为 BiLSTM 输出门通过线性变换得到。

$F(x)$ 计算如式(8)所示:

$$F(x) = W_{hf} o_t \quad (8)$$

其中: W_{hf} 为随机初始化权重矩阵。

1.3 句袋内注意力机制

$S^i \in R^{m_i \times 3d}$ 表示句袋 b^i 中所有句子表示, $R \in R^{h \times 3d}$ 表示关系嵌入矩阵,其中 h 是关系数量。与传统方法不同,传统方法推导了关系分类的统一句袋表示,本文方法在所有可能的关系条件下计算句袋 b^i 的表示 b_k^i :

$$b_k^i = \sum_{j=1}^m \alpha_{kj}^i S_j^i \quad (9)$$

其中: $k \in \{1, 2, \dots, k\}$ 是关系索引; α_{kj}^i 是句袋 b^i 中第 j 个句子和第 k 个关系的注意力权重。 α_{kj}^i 进一步定义如式(10)所示:

$$\alpha_{kj}^i = \frac{\exp(e_{kj}^i)}{\sum_{j'} \exp(e_{kj'}^i)} \quad (10)$$

其中: e_{kj}^i 是第 k 个关系查询与句袋 b^i 中第 j 个句子之间的匹配度,采用向量之间的简单点积^[15]来计算匹配度。 e_{kj}^i 计算如式(11)所示:

$$e_{kj}^i = r_k s_j^{iT} \quad (11)$$

其中: r_k 是关系嵌入矩阵 R_2 的第 k 行; T 是训练样本集合。

最终句袋 b^i 表示为矩阵 $B^i \in R^{h \times 3d}$,每行对应于此句袋中可能的关系类型。

1.4 句袋间注意力机制

为解决句袋带噪问题,设计一种基于相似性的

句袋间注意力模块^[16]来动态地降低带噪句袋的权重。如果两个句袋 b^{i1} 和 b^{i2} 都被标记为关系 k ,则 b_k^{i1} 和 b_k^{i2} 应该关系更接近,给定一组具有相同关系标签的句袋,将更高的权重分配给该组中与其他句袋接近的句袋,句袋组 g 可表述如下:

$$g_k = \sum_{i=1}^n \beta_{ik} b_k^i \quad (12)$$

其中: $g \in R^{h \times 3d}$; β_{ik} 组成注意力权重矩阵 $\beta \in R^{n \times k}$ 。 β_{ik} 计算如式(13)所示:

$$\beta_{ik} = \frac{\exp(\gamma_{ik})}{\sum_{i'=1}^n \exp(\gamma_{i'k})} \quad (13)$$

其中: γ_{ik} 表示用第 k 个关系标记句袋 b^i 的置信度^[17]。受到自注意力机制的启发^[18], γ_{ik} 使用向量本身计算一组向量的注意力权重,根据它们自身表示计算句袋的权重。 γ_{ik} 计算如式(14)所示:

$$\gamma_{ik} = \sum_{i'=1, 2, \dots, i' \neq i} \text{similarity}(b_k^i, b_k^{i'}) \quad (14)$$

函数相似性计算如式(15)所示:

$$\text{similarity}(b_k^i, b_k^{i'}) = b_k^i b_k^{iT'} \quad (15)$$

在计算式(12)~式(15)前,首先所有句袋表示 b_k^i 均标准化^[19]为单位长度,如式(16)所示:

$$\bar{b}_k^i = \frac{b_k^i}{\|b_k^i\|_2} \quad (16)$$

然后第 k 个关系得分 o_k 通过句袋组 g 和关系嵌入 r_k 计算得到,如式(17)所示:

$$o_k = r_k g_k^T + d_k \quad (17)$$

其中: d_k 是偏置项。

最后使用 Softmax 函数获得句袋组 g 被分类为第 k 个关系的概率,如式(18)所示:

$$p(k|g) = \frac{\exp(o_k)}{\sum_{k'=1}^h \exp(o_{k'})} \quad (18)$$

需要注意的是,相同的关系嵌入矩阵 R 用于计算式(11)和式(16),类似的 dropout 损失率用于句袋表示 B^i 以防止过拟合。

1.5 模型实现过程

基于 ResNet_BiLSTM 与句袋内和句袋间注意力机制的关系抽取模型实现过程具体如下:

1) 数据处理。首先训练集中的所有句子包含相同的两个实体,将其累加到一个句袋,然后对于每 n 个共用的句袋,将相同的关系标签放入一个句袋中,需要注意的是,一个句袋组是一个训练样本。因此,该模型也可以在小批量模式下通过打包多个句袋组成一批句袋。

2) 目标函数优化。优化公式如式(19)所示:

$$J(\theta) = - \sum_{(g, k) \in T} \log_a p(k|g; \theta) \quad (19)$$

其中: θ 是模型参数集,包括单词嵌入矩阵、位置特征嵌入矩阵、权重矩阵和关系嵌入矩阵; $J(\theta)$ 通过 mini-batch 随机梯度下降法 (Stochastic Gradient Descent, SGD) 来优化模型^[20]。

3)训练和测试。在训练阶段,将具有相同关系标签的 n 个句袋累积到 1 个句袋组中,并计算句袋表示的加权和,以获得句袋组 g 的表示。由于每个句袋的标签在测试阶段是未知的,因此在处理测试组时,每个句袋被视为一个句袋组(即 $n=1$)。此外,与文献[15]类似,仅对正样本应用句袋间注意力机制,其原因是表示无关系的句袋表示形式是多样的,难以计算权重。

4)预训练。在实验中采用预训练策略,首先对模型进行句袋内训练,直到收敛,然后添加句袋间注意力机制模块,进一步更新模型参数,直至再一次收敛。初步的实验结果表明,预训练策略相比于句袋间注意力机制能够获得更好的模型性能。

2 实验设置与结果分析

2.1 数据集和评价指标选取

实验采用 NYT(New York Times)数据集。该数据集由文献[21]发布并得到广泛使用,基于远程监督关系提取研究,将 Freebase 知识库中的三元组和 NYT 数据集中的文本对齐生成,包含 52 个实际关系和 1 个特殊关系 NA,其中 NA 表明 2 个实体之间没有关联性。

在计算机上使用 NVIDIA GTX 1080 Ti 显卡运行程序,采用精确率-召回率(Precision-Recall, PR)、曲线下面积(Area Under the Curve, AUC)和精确率(Precision, P)@ N (P@ N)^[22]来评估模型性能。P@ N 采用 One、Two 和 All 测试集,其中,One 表示对测试集中每个实体对随机选择一个句子,通过这一个句子对关系进行预测,Two 表示对测试集中每个实体对随机选择两个句子,通过这两个句子对关系进行预测,All 表示对测试集中每个实体对选择所有句子对关系进行预测,mean 表示对求得的结果取平均值。P@ N 使用了前 N 个实例的准确率,其中 N 取 100、200、300。

2.2 训练细节和超参数设置

在实验中,使用的多数超参数遵循文献[23]中的设置,如表 1 所示。在初始化时采用文献[19]发布的 50 维单词嵌入。2 个不同批量大小 N_p 和 N_t 分别用于预训练和训练,使用训练集进行网格搜索来确定 n 的最佳值, $n \in \{3, 4, \dots, 10\}$, $N_t \in \{3, 4, \dots, 200\}$, $N_p \in \{10, 20, 50, 100, 200\}$ 。需要注意的是,增加句袋组数量 n 可能会增强句袋间注意力,导致训练样本减

少,当 $n=1$ 时,句袋间注意力的影响将消失。使用 mini-batch SGD 的初始学习率为 0.1,学习率在每 100 000 步后下降至之前的 1/10,在实验中仅包含句袋内注意力的预训练模型,在 300 000 步内收敛,包含句袋间注意力的训练模型的初始学习率设置为 0.001。

表 1 实验超参数设置

Table 1 Setting of experimental superparameters	
参数名	参数值
词嵌入维度	50
位置特征最大相对距离	± 30
位置向量维度	5
损失率	0.5
优化函数	SGD
学习率	0.1
预训练样本数量 N_p	20
训练样本数量 N_t	10
句袋数量 n	5
隐藏层维度	230

2.3 模型性能对比

选取文献[24]中的 11 种模型与本文 ResNet_BiLSTM+ATT_RA+BAG_ATT 模型进行性能对比,其中,CNN、PCNN 和 ResNet_BiLSTM 分别表示不同句子编码方式,ATT_BL 表示基于句袋内注意力,ATT_RA 表示基于关系感知的句袋内注意力机制,BAG_ATT 表示基于句袋间注意力。在训练阶段,用于计算注意力权重的关系查询向量被固定为与每个句袋的远程监督标签相关联的嵌入向量^[25-26]。在测试阶段,所有关系查询向量都被应用于分别计算关系的后验概率,选择概率高的结果作为分类结果,给出所有模型的 AUC 值的平均值和标准差如表 2 所示。

表 2 不同模型的 AUC 值比较

Table 2 Comparison of AUC values of different models	
模型	AUC 值(平均值 \pm 标准差)
CNN+ATT_BL	0.376 \pm 0.003
CNN+ATT_RA	0.398 \pm 0.004
CNN+ATT_BL+BAG_ATT	0.388 \pm 0.002
CNN+ATT_RA+BAG_ATT	0.407 \pm 0.004
PCNN+ATT_BL	0.388 \pm 0.004
PCNN+ATT_RA	0.403 \pm 0.003
PCNN+ATT_BL+BAG_ATT	0.403 \pm 0.002
PCNN+ATT_RA+BAG_ATT	0.422 \pm 0.004
ResNet_BiLSTM+ATT_BL	0.391 \pm 0.002
ResNet_BiLSTM+ATT_RA	0.406 \pm 0.003
ResNet_BiLSTM+ATT_BL+BAG_ATT	0.405 \pm 0.003
ResNet_BiLSTM+ATT_RA+BAG_ATT	0.432 \pm 0.004

为进行定量比较,还绘制了所有模型的PR曲线图如图2~图4所示。由图2~图4可以看出:

1)ResNet_BiLSTM 作为句子编码器相比于 CNN、PCNN 表现更好。

2)使用 ResNet_BiLSTM、CNN 或 PCNN 作为句子编码器时,ATT_RA 优于 ATT_BL,主要因为 ATT_BL 在训练时推导出句袋表示时仅考虑目标关系,而 ATT_RA 以所有关系嵌入作为查询,计算出句袋内的注意力权重,提高了句袋表示的灵活性。

3)对于3种句子编码器和2种句袋内注意力机制,带有句袋注意力机制的模型相比于其他模型具有更好的性能,这一结果验证了句袋间注意力机制用于远程监督关系提取的有效性。

可见,将 ResNet_BiLSTM 作为句子编码器并与句袋内和句袋间注意力机制相结合可获得最佳 AUC 性能。

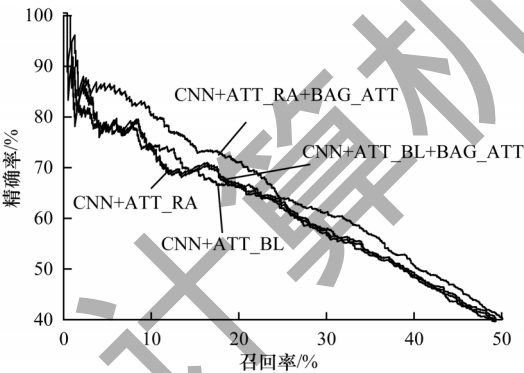


图2 CNN句子编码的PR曲线
Fig.2 PR curves of CNN sentence coding

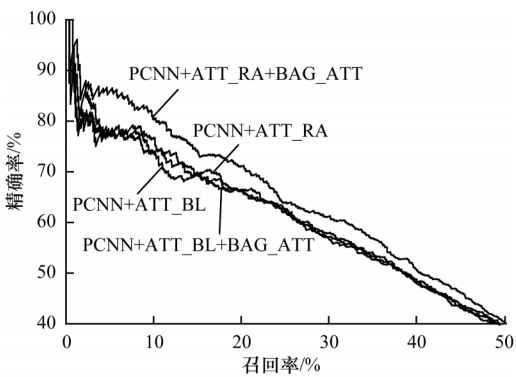


图3 PCNN句子编码的PR曲线
Fig.3 PR curves of PCNN sentence coding

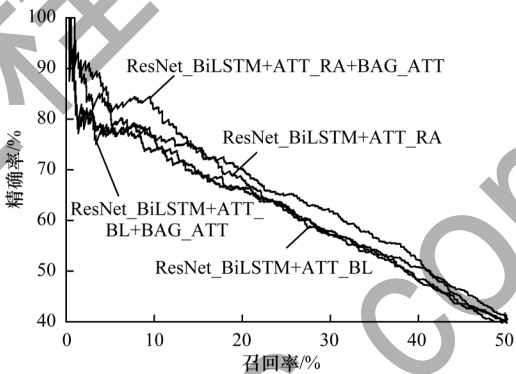


图4 ResNet_BiLSTM句子编码的PR曲线
Fig.4 PR curves of ResNet_BiLSTM sentence coding

2.4 句袋内注意力机制对模型性能的影响

通过实验验证句袋内注意力机制对模型性能的影响,随机选择实体对的1句、2句和所有句子进行测试并构造 One、Two、All 这3个测试集,实验结果如表3所示。

表3 在3种测试集上的模型P@N比较

Table 3 Comparison of P@N of models on three test sets

模型	One 测试集				Two 测试集				All 测试集				%
	P@100	P@200	P@300	mean	P@100	P@200	P@300	mean	P@100	P@200	P@300	mean	
文献[21]模型	73.3	69.2	60.8	67.8	77.2	71.6	66.1	71.6	76.2	73.1	67.4	72.2	
文献[20]模型	84.0	75.5	68.3	75.9	86.0	77.0	73.3	78.8	87.0	84.5	77.0	82.8	
CNN+ATT_BL	74.2	68.9	65.3	69.5	77.8	71.5	68.1	72.5	79.2	74.9	70.3	74.8	
CNN+ATT_RA	76.8	72.7	67.9	72.5	79.6	73.9	70.7	74.7	81.4	76.3	72.5	76.8	
CNN+ATT_BL+BAG_ATT	78.6	74.7	69.7	74.2	82.4	76.2	72.1	76.9	83.0	78.0	74.0	78.3	
CNN+ATT_RA+BAG_ATT	79.8	75.3	71.0	75.4	83.2	76.5	72.1	77.3	87.2	78.7	74.9	80.3	
PCNN+ATT_BL	78.6	73.5	68.1	73.4	77.8	75.1	70.3	74.4	80.8	77.5	72.3	76.9	
PCNN+ATT_RA	79.4	73.9	69.9	74.3	82.2	77.6	72.4	77.4	84.2	79.9	73.0	79.0	
PCNN+ATT_BL+BAG_ATT	85.2	78.2	71.3	78.2	84.8	80.0	74.3	79.7	88.8	83.7	77.4	83.9	
PCNN+ATT_RA+BAG_ATT	86.8	77.6	73.9	79.4	91.2	79.2	75.4	81.9	91.8	84.0	78.7	84.8	
ResNet_BiLSTM+ATT_BL	79.4	74.6	69.4	74.5	78.8	76.5	70.6	75.3	82.1	79.6	73.6	78.4	
ResNet_BiLSTM+ATT_RA	80.1	75.1	70.5	75.2	82.7	78.9	73.9	78.5	85.9	81.3	74.5	80.6	
ResNet_BiLSTM+ATT_BL+BAG_ATT	86.4	80.1	72.6	79.7	85.6	81.3	75.8	79.9	89.7	85.1	78.9	84.6	
ResNet_BiLSTM+ATT_RA+BAG_ATT	87.7	80.3	75.0	81.0	91.9	80.6	76.9	83.1	92.9	85.3	80.3	86.2	

由表3可以看出,ResNet_BiLSTM+ATT_AL+BAG_ATT具有较高的P@N值,无论采用ResNet_BiLSTM还是BAG_ATT,ATT_RA在所有实体对测试集上均优于ATT_BL。由于当一个句袋中只有一个句子时,ATT_BL和ATT_RA的解码程序是相同的,因此从ATT_BL到ATT_RA的改进可归因于ATT_RA在训练阶段以关系感知的方式计算句袋内注意力权重。

2.5 句袋间注意力权重分布

将句袋中的句子数设置为5进行句袋间注意力计算,每个句袋首先使用BAG_ATT模型来计算句袋间注意力机制的权重,然后计算训练集各部分句袋间注意力权重的平均值和标准差,如表4所示。由表4可以看出,训练句子数量较少的句袋通常被分配较低的句袋间注意力权重,且训练句子数量较少的实体对更可能有不正确的关系标签。

表4 不同句子数的句袋间注意力权重分布
Table 4 Distribution of attention weight between inter-sentence bags with different number of sentences

句子数	权重(平均值±标准差)
1	0.163 ± 0.029
2	0.187 ± 0.033
3	0.210 ± 0.034
4	0.212 ± 0.037
≥5	0.256 ± 0.043

3 结束语

本文提出基于ResNet_BiLSTM与句袋内和句袋间注意力机制的实体关系抽取模型。引入残差连接采集句子特征信息,保留句子在前后传递过程中的特征信息。通过BiLSTM进行句子特征信息识别与训练,解决了反向特征依赖问题。使用句袋内注意力机制,对正向实体与关系赋予更高权重,提高识别准确性。采用句袋间注意力机制,缓解了在提取句子中的关系时的噪声句袋问题。在NYT数据集上的实验结果表明,该模型能够充分利用实体与关系特征,具有更好的关系抽取性能。下一步将使用实体关系与句子的分区过滤策略,并结合BERT提取语义特征,进一步提高关系抽取的准确性和灵活性。

参考文献

- [1] HAN S Q, HAO X L, HUANG H L. An event-extraction approach for business analysis from online Chinese news[J]. Electronic Commerce Research and Applications, 2018, 28: 244-260.
- [2] FENG J, HUANG M L, ZHAO L, et al. Reinforcement learning for relation classification from noisy data[C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2018: 1-9.
- [3] ZELENKO D, AONE C, RICHARDELLA A. Kernel methods for relation extraction[J]. Journal of Machine Learning Research, 2020, 3(3): 1083-1106.
- [4] 孙紫阳,顾君忠,杨静. 基于深度学习的中文实体关系抽取方法[J]. 计算机工程, 2018, 44(9): 164-170.
SUN Z Y, GU J Z, YANG J. Chinese entity relation extraction method based on deep learning[J]. Computer Engineering, 2018, 44(9): 164-170. (in Chinese)
- [5] 鄂海红,张文静,肖思琪,等. 深度学习实体关系抽取研究综述[J]. 软件学报, 2019, 30(6): 1793-1818.
E H H, ZHANG W J, XIAO S Q, et al. Survey of entity relationship extraction based on deep learning[J]. Journal of Software, 2019, 30(6): 1793-1818. (in Chinese)
- [6] BUNESCU C, MOONEY R. A shortest path dependency kernel for relation extraction[C]//Proceedings of Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Philadelphia, USA: Association for Computational Linguistics, 2005: 724-731.
- [7] KAMBHATLA N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations[C]//Proceedings of the 43rd ACL International Conference. Philadelphia, USA: Association for the Computational Linguistics, 2017: 178-181.
- [8] UDDIN J, GHAZALI R, DERIS M M. Does number of clusters effect the purity and entropy of clustering[C]// Proceedings of 2016 International Conference on Soft Computing and Data Mining. Berlin, Germany: Springer, 2016: 355-365.
- [9] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [10] CHEN C M, GUAN D J, SU Q K. Feature set identification for detecting suspicious URLs using Bayesian classification in social networks[J]. Information Sciences, 2014, 289: 133-147.
- [11] ZHANG M, ZHANG J, SU J, et al. A composite kernel to extract relations between entities with both flat and structured features[C]//Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2006: 825-832.
- [12] QIN P D, XU W R, WANG W Y. Robust distant supervision relation extraction via deep reinforcement learning[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2018: 203-209.
- [13] LI L F, NIE Y P, HAN W H, et al. A multi-attention-based bidirectional long short-term memory network for relation extraction[C]//Proceedings of International Conference on Neural Information Processing. Berlin, Germany: Springer, 2017: 216-227.
- [14] CASO A, ROSSI S. Users ranking in online social networks to support POI selection in small groups[C]//Proceedings of International Conference on Neural Information Processing. Berlin, Germany: Springer, 2020: 104-107.
- [15] GAO J, ZHOU T. Evaluating user reputation in online rating systems via an iterative group-based ranking method[J]. Physica A: Statistical Mechanics and Its Applications, 2017, 473: 546-560.

(上接第 115 页)

- [16] PENG N Y, POON H, QUIRK C, et al. Cross-sentence N-ary relation extraction with graph LSTMs[J]. Transactions of the Association for Computational Linguistics, 2017, 5: 101-115.
- [17] HUANG YI Y, WANG W Y. Deep residual learning for weakly supervised relation extraction [J]. Journal of Machine Learning Research, 2019, 4(2): 103-106.
- [18] DIETTERICH T G, LATHROP R H, LOZANO-PÉREZ T. Solving the multiple instance problem with axis-parallel rectangles[J]. Artificial Intelligence, 1997, 89(1/2): 31-71.
- [19] HAO Y C, ZHANG Y Z, LIU K, et al. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2017: 221-231.
- [20] LIU H C, YOU J X, LI Z W, et al. Fuzzy Petri nets for knowledge representation and reasoning: a literature review[J]. Engineering Applications of Artificial Intelligence, 2017, 60(C): 45-56.
- [21] LIN Y K, SHEN S Q, LIU Z Y, et al. Neural relation extraction with selective attention over instances[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2016: 2124-2133.
- [22] 刘正铭, 马宏, 刘树新, 等. 一种融合节点文本属性信息的网络表示学习算法[J]. 计算机工程, 2018, 44(11): 165-171.
- LIU Z M, MA H, LIU S X, et al. A network representation learning algorithm fusing with textual attribute information of nodes[J]. Computer Engineering, 2018, 44(11): 165-171. (in Chinese)
- [23] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [24] YE Z X, LING Z H. Distant supervision relation extraction with intra-bag and inter-bag attentions[EB/OL]. [2021-09-11]. <https://arxiv.org/pdf/1904.00143.pdf>.
- [25] YAN X, MOU L L, LI G, et al. Classifying relations via long short term memory networks along shortest dependency path[EB/OL]. [2021-09-11]. <https://arxiv.org/abs/1508.03720>.
- [26] KILINC O, UYSAL I. GAR: an efficient and scalable graph-based activity regularization for semi-supervised learning [J]. Neurocomputing, 2018, 296: 46-54.

编辑 陆燕菲