

基于多层次注意力机制的远程监督关系抽取模型

李 浩 刘永坚 解 庆 唐伶俐

(武汉理工大学计算机科学与技术学院 武汉 430070)

(国家新闻出版署出版融合发展(武汉)重点实验室 武汉 430070)

摘 要 实体关系抽取作为信息抽取的主要任务之一,其目的在于确定无结构文本中两个实体的关系类别。目前准确率较高的有监督方法由于需要大量的人工标注语料而受到了限制,而远程监督方法则通过知识库与文本集进行启发式对齐来获取大量关系三元组,这是解决大规模关系抽取任务的主要途径。针对目前远程监督关系抽取的研究未能充分利用句子上下文词语的高层语义,以及未考虑关系之间的依赖包含关系的问题,文中提出了一种基于多层次注意力机制的远程监督关系抽取模型。该模型首先通过双向 GRU(Gate Recurrent Unit)神经网络对句子词向量进行编码来获取句子高维语义;其次通过引入词语层注意力来计算两个实体与上下文词语的相关程度,从而充分捕捉句子中实体上下文的语义信息;然后在多个实例上构建句子层的注意力来减少标签错误标注的问题;最后通过关系层的注意力自动学习不同关系之间的依赖包含关系。在 FreeBase+NYT 公共数据集上的实验结果表明,在双向 GRU 模型的基础上引入词语层、句子层和关系层注意力机制对提高远程监督关系抽取的效果都起到了促进作用;将三层注意力机制进行融合得到的多层次注意力机制关系抽取模型的准确率和召回率相较于现有的主流方法提高了 4% 左右,更好地实现了关系抽取,从而为进一步构建知识图谱、智能问答等应用奠定了理论基础。

关键词 远程监督,关系抽取,双向 GRU,词向量,注意力机制

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/jsjx.180901780

Distant Supervision Relation Extraction Model Based on Multi-level Attention Mechanism

LI Hao LIU Yong-jian XIE Qing TANG Ling-li

(School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China)

(State Press and Publication Administration Publishing Fusion Development Key Laboratory, Wuhan 430070, China)

Abstract As one of the main tasks of information extraction, entity relation extraction aims at determining the relationship category of two entities in unstructured text. At present, the supervised method with high accuracy is limited by the need for a large number of manual tagging corpus. The distant supervision method obtains a large number of relational triples by heuristic alignment between knowledge base and text set, which is the main way to solve the large-scale relational extraction task. In order to solve the problems that the high-dimensional semantics of words in sentence context are not fully utilized and the dependency-inclusion relationship between relationships is not considered in the current research on distant supervision relation extraction, this paper proposed a multi-level attention mechanism model for distant supervision relation extraction. In this model, the high-level semantics of sentences are obtained by utilizing the bidirectional GRU(Gate Recurrent Unit) neural network to code the sentence word vectors. Then, the word-level attention is introduced to calculate the degree of correlation between two entities and the context words, thus capturing the semantic information of the entity context in sentences adequately. Next, the sentence-level attention is constructed on multiple instances to reduce the tag error annotation problem. Finally, the dependency-inclusion relationship between different relationships is automatically learned by the relation-level attention. The experimental results on FreeBase+NYT public dataset show that the introduction of word-level, sentence-level and relation-level attention mechanisms on the basis of bidirectional GRU model can improve the effect of distant supervision relation extraction. Compared with the existing mainstream methods, the multi-level attention mechanism relation extraction model obtained by integrating three levels attention mechanisms improves the accuracy and recall rate by about 4%, which achieves better relation extraction effect, thus providing a theoretical foundation for further constructing the knowledge graph and intelligent question answering applications.

Keywords Distant supervision, Relation extraction, Bidirectional GRU, Word embedding, Attention mechanism

到稿日期:2018-09-20 返修日期:2018-12-19 本文受国家自然科学基金(61602353),湖北省自然科学基金(2017CFB505)资助。

李 浩(1992-),男,硕士生,主要研究方向为知识服务和自然语言处理;刘永坚(1962-),男,博士,教授,主要研究方向为文化资源数字化、数字出版和数据传播;解 庆(1986-),男,博士,副教授,CCF 会员,主要研究方向为流数据挖掘、知识服务和推荐系统,E-mail: felixxq@whut.edu.cn(通信作者);唐伶俐(1989-),女,博士,讲师,主要研究方向为数字传播和版权保护。

近年来,互联网信息的爆炸式增长给人们快速而精确地获取需要的信息带来了巨大的困难,基于关键词匹配和 PageRank 等网页排序技术的搜索引擎虽然能够在一定程度上满足用户信息获取的需求,但仍存在信息过载、资源迷向等问题。信息抽取技术能够对文本进行深层的分析,从而为用户提供更准确的信息服务。实体关系抽取作为信息抽取的主要任务,其主要目的是从无结构化的纯文本中提取出二元关系,并以三元组〈实体 e1, 实体 e2, 关系 r〉的形式进行表示^[1]。例如,给定一个句子:“〈e1〉Trump〈/e1〉 was born in the 〈e2〉New York 〈/e2〉 just as he has always said.”,通过关系抽取将自动识别这个句子的关系:〈Trump, New York, Born In〉,表示“Trump”和“New York”的关系是“Born In”。因此,关系抽取对于知识图谱的构建、信息检索、智能问答等领域有着广泛的应用价值。

目前,实体关系抽取最常用的方法是有监督的关系抽取,尤其将深度学习框架的神经网络模型应用到实体关系抽取任务上,关系抽取的效果更加明显。虽然有监督的关系抽取的准确率高,但该方法需要大量的人工标注语料,耗时耗力,限制了大规模的实体关系抽取。针对标注语料不足的问题,Mintz 等^[2]提出了一种基于知识库与大规模文本启发式对齐的远程监督关系抽取方法,以获取关系三元组,但是远程监督方法很容易引入噪声标注数据,从而导致关系抽取效果不佳。为了提高远程监督关系抽取的性能和效果,本文提出了一种基于多层次注意力机制的远程监督关系抽取模型,分别对词语、句子和关系层采用了注意力机制:首先通过词语层的注意力机制来充分捕捉句子上下文的语义信息;然后通过句子和关系之间句子层的注意力机制来减少错误标注问题;最后通过关系层的注意力机制来学习关系之间的依赖关系。实验结果表明,与当前具有代表性的方法进行比较,本文提出的模型在关系抽取任务中的效果得到了显著的提高。

1 相关工作

在众多关系抽取方法中,有监督的关系抽取方法是最常见的基本方法,其主要思想是在已经标注的训练数据集上训练出关系分类模型,然后在测试数据集上进行关系类型识别。目前的研究主要分为基于特征向量的方法^[3]、基于核函数的方法^[4]和基于神经网络模型的深度学习的方法^[5]。基于特征向量的方法结合启发式方法来提取特征集合,以构建特征向量,采用合适的分类器构建抽取模型,然而语法的复杂结构和大量的人工特征标记使得基于特征工程的方法性能难以继续提升。基于核函数的方法不需要构造复杂的特征工程,而是利用语料本身的结构信息,通过核函数计算两个非线性结构的相似度,从而进行实体关系抽取。

近年来,深度学习框架的神经网络模型被广泛应用到实体关系抽取任务上,相比于传统的基于特征和核函数的方法,神经网络模型能自动学习句子特征,从而避免了特征选择和 NLP 工具带来的错误传播等问题^[6],同时其关系抽取的效果也有了很大的提高。虽然有监督的关系抽取的准确率高,但该方法需要大量的人工标注语料,限制了大规模的实体关系抽取。

为了解决人工标注语料严重缺乏的问题,Mintz 等^[2]提出了一种不需要人工标注的远程监督方法,即利用 FreeBase

知识库和 Wikipedia 文本进行对齐,从而获取大规模关系三元组。远程监督基于这样一个假设:如果两个实体包含某种关系,那么所有包含这两个实体的句子都表达这种关系。由于这种假设过于严格,很容易引起噪声标注问题,因此噪声标注的过滤成为了远程监督方法的研究重点。Hoffmann 等^[7]和 Surdeanu 等^[8]分别提出了多实例学习和多实例多标签的方法来缓解错误标注问题;Benjamin 等^[9]则通过结合主题模型 LDA 和判决学习方法来减少远程监督结果中的噪声数据。近年来,不少研究也开始将深度学习应用到远程监督关系抽取任务中,Zeng 等^[10]采用分段最大池的卷积网络自动学习句子特征,并结合多实例学习进行远程监督关系抽取;Jiang 等^[11]提出了一种多实例多标签的卷积神经网络远程监督方法,该方法充分利用了不同句子之间的共享信息,从而实现了较好的抽取效果;Lei 等^[12]为了有效地模拟文本语料库与知识图谱信息之间的关系模式,提出了一种具有双向知识蒸馏的神经关系抽取框架,该框架可以协同使用不同的信息源,减少远程监督中的噪声标签问题。

注意力机制最早是由 Treisman 和 Gelade 提出的一种人脑注意力模型,通过计算概率分布突出关键输入信息对输出的影响,并以此来优化模型。在自然语言处理领域,Bahdanau 等^[13]首次采用注意力机制对机器翻译模型进行优化,并得到很好的优化效果。Lin 等^[14]在 Zeng 等^[10]的远程监督关系抽取模型中引入了句子层注意力机制,并充分利用语料中所有包含两个实体对的句子,通过赋予关系标注正确实例更高的权重,来减少噪声标注。Ji 等^[15]在考虑句子层的注意力机制的同时,从知识库中引入了已标注实体的描述信息,加强了对实体的向量表示学习,从而提高了关系抽取准确率。为了解决长句子距离依赖以及语料数据假阳性问题,He 等^[16]提出了基于单词和句子层注意力的长短期记忆神经网络的远程监督关系抽取模型,得到了比传统方法更好的抽取效果和性能。

虽然采用注意力机制优化远程监督关系抽取模型取得了比传统方法更好的分类效果,但目前大多数研究仅在句子层引入注意力机制,尚未考虑句子中不同词语对实体对的重要程度,以及关系之间的相互依赖包含关系。在生成句子向量时,句子中的词语对实体对的影响程度是不同的,例如,对于在“〈e1〉Trump〈/e1〉 was born in the 〈e2〉New York 〈/e2〉 just as he has always said.”中,单词“born”对关系三元组〈Trump, New York, Born In〉的重要程度大于单词“said”;同时,不同的关系之间存在着依赖包含关系,例如,对于三元组〈Beijing, China, Capital Of〉和三元组〈Beijing, China, Part Of〉,通过简单的隐含关系推断可以判定这两种关系属于同一类。因此,为了充分结合注意力机制并提高关系抽取的效果,本文在采用句子层注意力机制的基础上,通过词语层注意力来充分捕捉句子上下文的语义信息,并通过关系层的注意力来自动学习关系之间的依赖包含关系。

2 多层次注意力机制的关系抽取模型

本文提出了一种多层次注意力机制的远程监督关系抽取模型,该模型结构如图 1 所示。首先将输入句子中所有词语进行词向量表示,并采用双向门控循环网络单元(Bidirectional Gate Recurrent Unit, BGRU)对词向量进行编码以获取句子的高维语义信息;然后通过计算上下文词语与实体对的相关

程度,建立词语层的注意力机制来充分捕捉句子上下文的语义信息;通过最大池化获取句子的向量表示,接着通过计算句子与关系的相关程度来建立句子层注意力机制,减少标签的

错误标注问题;同时,引入关系层的注意力机制来自动学习不同关系之间的相互依赖包含关系,最后通过非线性分类器将关系表示进行分类,从而获取最终的关系抽取结果。

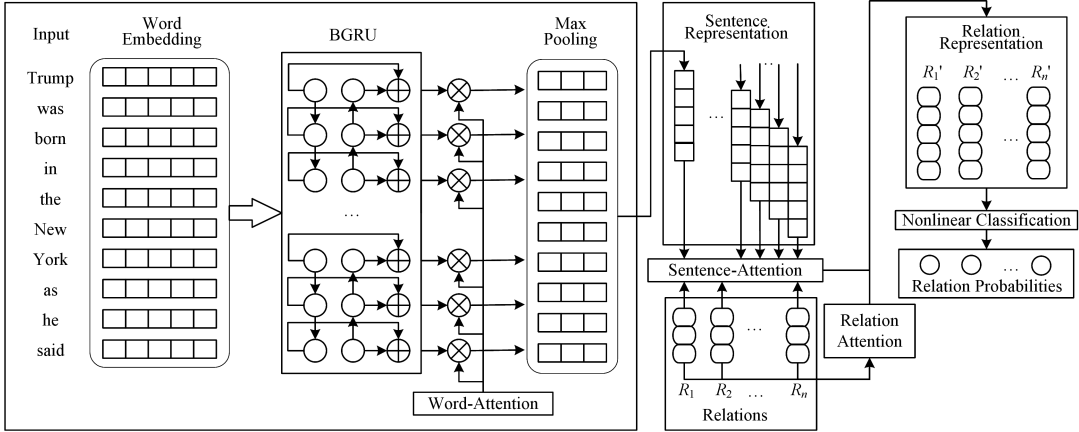


图1 多层次注意力机制的远程监督关系抽取模型

Fig.1 Distant supervision relation extraction model of multi-level attention mechanism

2.1 词向量表示

构建词向量是将句子文本转换为向量形式,将每个句子词语转换为一个多维向量矩阵^[17]。给定一个含有 m 个词语的句子 $S=(w_1, w_2, \dots, w_m)$,每个词语都需要经过向量化处理,并使用词向量进行表示:

$$w_i = W^{word} \times V^w \tag{1}$$

其中, $W^{word} \in R^{l \times |m|}$ 表示采用 word2vec^[18] 训练文本的词向量矩阵, l 表示词向量的维数 ($l=300$), m 表示句子词语个数, V^w 是词语的 one-hot 表示。

为了能够充分获取句子中词语的句法和语义信息,本文还将采用实体相对位置特征对词向量进行强化表示。

相对位置特征(PF)是指句子中词语相对于两个实体的相对距离,例如“ $\langle e1 \rangle$ Trump $\langle /e1 \rangle$ was born in the $\langle e2 \rangle$ New York $\langle /e2 \rangle$ just as he has always said.”中,单词“born”到头实体“Trump”的相对距离为 2,到尾实体“New York”的相对距离为 -3。将这两个相对距离进行向量表示,形成实体的相对位置向量 w_i^e 和 w_i^b 。

最后将词语的词向量与实体相对位置向量进行级联,得到完整的词向量表示。

2.2 BiGRU 编码

为了能够充分获取句子的高维语义表示,并解决句子的长距离依赖问题,本文采用双向 GRU 模型对输入的词向量进行编码。GRU 模型是 Chung 等^[19]提出的一种 LSTM 模型的变种,既保持了 LSTM 的训练效果,又简化了 LSTM 的内部结构,而且提高了模型的训练速率。它包含两个门结构(更新门和重置门)和一个隐藏状态,其内部结构如图 2 所示。

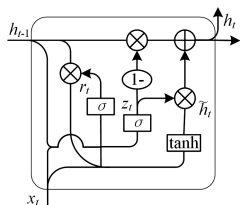


图2 GRU 模型的内部结构

Fig.2 Internal structure of GRU model

将句子第 t 个词语的词向量 $x_t=(w_t, w_t^e, w_t^b)$ 作为 GRU 模型的输入,获取第 t 个词的 GRU 各个门单元状态的特征值,具体如下:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{2}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{3}$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t] + b_h) \tag{4}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{5}$$

其中, z_t 表示更新门, r_t 表示重置门, \tilde{h}_t 表示隐藏状态, h_t 表示输出状态; W 为权重矩阵, b 为偏置向量,这两个参数都是通过训练学习得到; $[\]$ 表示向量相连接, $*$ 表示矩阵元素相乘。

通过双向 GRU 编码可以获得第 t 个词语的输出状态,假设正向输出状态为 \vec{h}_t ,反向输出状态为 \overleftarrow{h}_t , $h_t = \vec{h}_t \oplus \overleftarrow{h}_t$ 则充分表达了句子单个词语的高维语义信息。

2.3 多层次注意力机制

2.3.1 引入词语层注意力

在关系抽取任务中,对于句子语义,并不是所有词语都有着相同的重要性,对于不同的实体对,上下文词语的重要程度也是不同的。例如,在句子“The United States President Trump will attend the NATO summit in Brussels, the capital of Belgium”中,词语“President”对于实体对 \langle United States, Trump \rangle 比词语“capital”更重要,相反,在实体对 \langle Brussels, Belgium \rangle 中,“capital”比“President”更重要。因此,本文通过计算上下文词语和实体对之间的相关程度,建立词语层的注意力机制来充分捕捉句子上下文的语义信息。

假设由 BiGRU 编码得到句子的向量表示 $H = \{h_1, h_2, \dots, h_m\}$ 和实体对表示 (h_{ch}, h_{ct}) ,通过计算句子中上下文词语与实体对的相关程度来获取句子的向量表示 x :

$$x = \sum_{i=1}^m \alpha_i h_i \tag{6}$$

其中, m 为句子长度, $\alpha_i \in [0, 1]$ 为上下文词语相对于实体对的权重,且 $\sum_i \alpha_i = 1$ 。这里使用一个前馈神经网络来计算词语与实体对的语义相关性,评分函数的计算公式如下:

$$\mu_i = \tanh(W_{word-att} [h_i, h_{ch}, h_{ct}] + b_{word-att}) \tag{7}$$

其中, $\mathbf{W}_{\text{word-att}}$ 为词语相关性权重矩阵, $\mathbf{b}_{\text{word-att}}$ 为偏置值。然后通过 softmax 函数将计算得到的 $\mu_1, \mu_2, \dots, \mu_m$ 进行归一化, 从而得到权重 α_i :

$$\alpha_i = \frac{\exp(\mu_i)}{\sum_k^m \exp(\mu_k)} \quad (8)$$

最后, 采用最大池化策略。选择句子的最显著特征, 得到句子表示:

$$s = \max \text{pool}(x) \quad (9)$$

2.3.2 引入句子层注意力

为了减少传统远程监督关系抽取带来的噪声标注问题, 本节采用与文献[14]相同的句子层注意力机制的多实例学习, 通过计算包含同一实体对的所有句子和预测关系的相关程度, 为与关系向量相关程度更高的句子赋予更高的权重, 从而降低噪声句子对关系预测的影响。

假设含相同实体对的 k 个句子组成的句子集为 $S = \{s_1, s_2, \dots, s_k\}$, 为了能够突出所有句子与预测关系的相关程度, 可以将句子向量表示为 S 中所有句子向量的加权和形式:

$$R = \sum_{i=1}^k \beta_i s_i \quad (10)$$

其中, $\beta_i \in [0, 1]$ 为句子集中每个句子相对于预测关系的权重, 且 $\sum_i \beta_i = 1$ 。通过计算句子向量和知识库中实体对应的关系向量的内积, 并对内积使用 softmax 函数进行归一化, 可以得到句子层的注意力权重 β_i :

$$\beta_i = \text{softmax}(s_i \mathbf{A} \mathbf{v}_r) \quad (11)$$

其中, \mathbf{A} 为随机初始化的加权对角矩阵, \mathbf{v}_r 为关系 r 的向量表示。

2.3.3 引入关系层注意力

假设存在一个集合 C , 其包含每个实体对的 n 个关系表示, 即 $C = \{R_1, R_2, \dots, R_n\}$ 。为了能够充分利用不同关系之间的相互依赖包含关系, 我们采用与句子层相同的注意力模型来计算每个关系之间的相似性, 关系 R_i 可以通过其他关系进行加权表示:

$$R_i^* = \sum_{j=1}^n \gamma_{ij} R_j \quad (12)$$

其中, $\gamma_{ij} \in [0, 1]$ 表示关系 R_i 与关系 R_j 之间的相似性权重, 且 $\sum_j \gamma_{ij} = 1$ 。通过关系 R_i 与关系 R_j 的向量内积计算 γ_{ij} , 即:

$$\gamma_{ij} = \text{softmax}(R_i \mathbf{B} R_j) \quad (13)$$

其中, \mathbf{B} 为加权对角矩阵。

2.4 模型训练与优化

将多层注意力机制输出的关系表示为 $\{R_1^*, R_2^*, \dots, R_n^*\}$, 通过线性表示得到每种关系的正确分类:

$$o_i = \mathbf{W}_i R_i^* + \mathbf{b}_i \quad (14)$$

其中, \mathbf{W}_i 是每个关系标签的权重矩阵, \mathbf{b}_i 为偏移向量。然后通过 sigmoid 逻辑函数预测每种关系的条件概率:

$$p(i|M, \theta) = \frac{1}{1 + \exp(-o_i)} \quad (15)$$

其中, M 为对齐句子集合, $i \in (1, 2, \dots, k)$, k 为关系的数量。

为了获取多标签分类的最优模型, 本文采用交叉熵损失函数作为目标函数, 定义如下:

$$J(\theta) = - \sum_{i=1}^k \log p(i|M, \theta) \quad (16)$$

其中, θ 表示模型可训练参数, 本文通过反向传播计算所有参数的梯度, 并使用 Adadelta 优化器进行参数更新。

为了防止出现过拟合现象, 并提高模型的训练速度, 本文还将采用 Dropout^[20] 对关系抽取模型进行正则化约束。Dropout 将主要用在双向 BGRU 编码层, 在保持输入、输出神经元不变的前提下, 随机地丢弃隐藏层的部分神经元。

3 实验与分析

3.1 数据集与评估标准

本文使用的数据集 FreeBase + NYT^[21] 是由 Riedel 于 2010 年通过将 Freebase 知识库中的实体对与纽约时报语料库(NYT)进行启发式对齐生成的, 近年来, 被广泛应用于远程监督关系抽取任务中。本文将 2005—2006 年语料库对齐句子作为训练数据, 将 2007 年语料库对齐句子作为测试数据。该数据集共包含特殊关系 NA(即实体对无关系)在内的 53 种关系; 训练集包含 522 611 个句子、281 270 个实体对、18 252 个关系事实; 测试集包含 172 448 个句子、966 78 个实体对、1950 个关系事实。

本文主要采用准确率-召回率(PR)曲线、F 值(F_score)和平均准确率(P@N)作为评价指标, 来评估提出的方法是否有效。准确率、召回率和 F 值的计算公式如下:

$$\text{precision} = \frac{\text{out_right}}{\text{out_all}} \quad (17)$$

$$\text{recall} = \frac{\text{out_right}}{\text{test_all}} \quad (18)$$

$$F_score = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (19)$$

其中, out_right 表示输出中预测正确的关系数量, out_all 表示输出中总共的关系数量, test_all 表示测试集中总共的关系数量。

3.2 实验环境与参数设置

实验环境: 操作系统 Windows10, 64 位; 处理器 Intel(R) Xeon(R) CPU E3-1225 v5@3.30GHz; 内存 16GB; 编程平台 Pycharm3.6。

本文采用交叉验证方法对模型参数进行调优, 在参数设置过程中, 参照文献[14]的经验对数据进行实验验证, 具体的参数设置如表 1 所列。

表 1 实验参数设置

Table 1 Setting of experimental parameters

参数	值
词向量维度 d_w	50
位置向量维度 d_{ab}	5
句子向量维度 d_s	230
关系向量维度 d_r	230
学习率 μ	0.001
批大小 B	60
Dropout 率 ρ	0.5

3.3 实验结果及分析

3.3.1 所提方法与现有方法的比较

本文将提出的方法与当前具有代表性的远程监督方法进行比较。

基于特征的传统远程监督方法有: 1) Mintz 方法, 即

Mintz 等首次提出的远程监督方法,采用多类逻辑回归分类器进行分类;2) Hoffmann 方法,即 Hoffmann 等提出的一种基于“at-least-one”确定性假设的多实例学习方法;3) MIML 方法,即 Surdean 等提出的一种采用图模型进行远程监督的多实例多标签方法。

基于神经网络的远程监督方法有:1) PCNN+MIL,即 Zeng 等采用的分段最大池的卷积神经网络并结合多实例学习的远程监督方法;2) PCNN+ATT,即 Liu 等在 Zeng 的基础上引入句子层注意力机制来减少噪声标注的方法。

将以上 5 种方法与本文提出的方法进行对比,绘制的准确率-召回率(PR)曲线如图 3 所示。

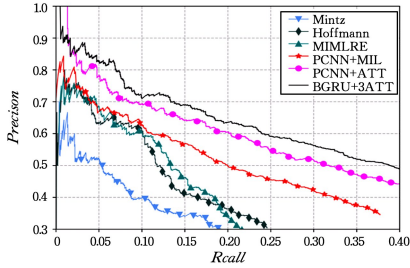


图 3 不同方法的 PR 曲线

Fig. 3 PR curve of different methods

从 PR 曲线可以看出,基于神经网络的方法明显优于基于特征的传统方法,主要原因是神经网络模型能够自动学习句子特征,从而避免了特征选择和 NLP 工具带来的错误传播问题;相较于其他神经网络方法,本文提出的多层次注意力机制方法有着较高的准确率和召回率,比 PCNN+ATT 模型提高了 4%~5%,这说明在 BGRU 模型上引入词语层、句子层和关系层注意力机制对提高远程监督关系抽取的效果有一定的促进作用。

3.3.2 不同层注意力机制对模型的影响

多层次注意力机制是该关系抽取模型的最重要部分,该实验部分将采用注意力增量的方法,分别在 BGRU 模型的基础上引入不同层次的注意力机制,从而验证词语层、句子层和关系层的注意力机制对模型的影响效果。同样使用准确率-召回率(PR)曲线进行评估,实验结果如图 4 所示。

从图 4 可以看出,在 BGRU 模型的基础上分别引入不同层次的注意力机制后的准确率,相较于单纯的 BGRU 关系抽取模型均有进一步的提升,从而验证了在模型中增加词语层、句子层和关系层注意力机制对关系抽取效果有一定的促进作用;同时,句子层注意力机制对模型的影响程度要比其他两种

注意力机制略高一点,主要原因是句子层的选择注意力能够充分考虑到包中所有句子的全局信息;采用三层注意力机制的 BGRU+3ATT 方法优于其他单层注意力机制方法,这说明 BGRU+3ATT 能够融合多层注意力机制,两者取长补短,相互促进,能够有效提高远程监督关系的抽取效果。

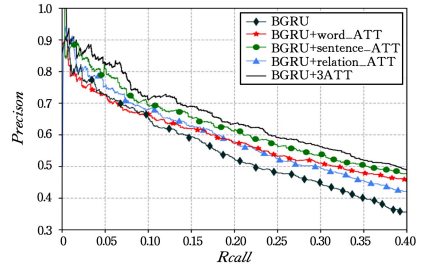


图 4 不同层次的注意力机制对模型的影响

Fig. 4 Effect of different level attention mechanism on model

为了进一步验证关系层注意力对语料句子中依赖包含关系的影响,实验选取了已知关系语料中具有依赖包含关系的两组种子关系 (/people/person/children 和 /people/family/members, 以及 /location/location/contain 和 /location/country/capital),然后采用 BGRU 和 BGRU+ATT_{relation} 模型进行对比实验,并采用 F 值对其进行性能评估,实验结果如表 2 所列。

表 2 关系层注意力对依赖关系的影响

Table 2 Effect of relation-level attention on dependency relations

关系	BGRU/%	BGRU+ATT _{relation} /%
/people/person/children	56.24	59.13
/people/family/members	54.91	58.68
/location/location/contain	59.60	65.21
/location/country/capital	61.35	64.46

从表 2 可以看出,引入关系层注意力提高了依赖包含关系的抽取效果,并且依赖关系都具有相似的 F 值,其主要原因是依赖包含关系具有相似的关系表示,这些相似的关系可以判断为同一类关系,这降低了关系分类过程中的误差。

3.3.3 不同句子数量的影响

在原始测试数据集中,有 74 857 个实体对仅对应一个句子,这几乎占了所有实体对的 3/4。由于句子层和关系层注意力机制的优势在于包含多个句子的实体对,因此本文将采用平均准确率 $P@N$ 来评估各模型对含有多个句子的实体对的关系抽取性能。这里,词语层注意力机制的目的在于获取句子的高层语义信息,与包含实体对的句子数量关系不大,因此不做对比。实验结果如表 3 所列。

表 3 包含实体对的不同句子数量的平均准确率 $P@N$

Table 3 Average accuracy $P@N$ of different sentences containing entity pairs

(单位:%)

测试集	One				Two				All			
	100	200	300	Mean	100	200	300	Mean	100	200	300	Mean
BGRU	69.9	66.7	61.1	65.9	71.4	69.6	61.8	67.6	72.2	68.8	64.4	68.5
BGRU+ATT _{sentence}	74.7	70.8	63.9	69.8	76.6	73.9	67.4	72.6	76.5	74.6	70.2	73.8
BGRU+ATT _{relation}	72.3	67.5	62.6	67.5	74.7	69.1	66.5	70.1	75.3	71.4	67.8	71.5
PCNN+ATT	73.3	69.2	60.8	67.8	77.2	71.6	66.1	71.6	76.2	73.1	67.4	72.2
BGRU+3ATT	79.2	73.1	66.4	72.9	80.3	75.9	71.6	75.9	82.6	77.7	72.5	77.6

由表 3 可以看出:1)相较于基准方法 PCNN+ATT,BG-

RU+3ATT 方法的 $P@N$ 值提高了 5%左右,再次验证了多

层次注意力机制对远程监督关系抽取的有效性;2)相较于单纯的 BGRU 方法,BGRU+ATT_{sentence} 和 BGRU+ATT_{relation} 方法的 $P@N$ 值都有了较明显的提高,证明了句子层和关系层注意力机制对多实例学习的促进作用,同时句子层注意力机制对关系抽取的促进效果也要高于关系层注意力机制。

结束语 本文提出了一种多层次注意力机制的远程监督关系抽取方法,分别在词语层、句子层以及关系层采用了注意力机制,词语层注意力机制能够充分获取句子上下文的高层语义信息,句子层注意力机制能够减少标签的错误标注问题,关系层注意力机制则能够自动学习不同关系之间的依赖包含关系。将 3 层注意力机制进行融合、协同训练,其在公共真实数据集上的实验结果表明,本文提出的方法相较于现有方法有了较大的进步。然而,使用 GRU 神经网络对句子进行编码,虽然能够解决句子的长距离依赖问题,但在捕获句子局部关键特征时,其性能不及 CNN 网络。在未来的工作中,我们会将 GRU 与 CNN 相结合来表征句子向量,以获取更加丰富的特征。

参 考 文 献

- [1] LIU S Y, LI B C, GUO Z Z, et al. Review of Entity Relation Extraction[J]. Journal of Information Engineering University, 2016, 17(5): 541-547. (in Chinese)
刘绍毓,李弼程,郭志刚,等. 实体关系抽取研究综述[J]. 信息工程大学学报, 2016, 17(5): 541-547.
- [2] MINTZ M, STEVEN B, RION S, et al. Distant supervision for relation extraction without labeled data [C]// Proceedings of Joint Conference of the Meeting of the ACL. Stroudsburg: Association for Computational Linguistics, 2009: 1003-1011.
- [3] GAN L X, WAN C X, LIU D X, et al. Chinese Named Entity Relation Extraction Based on Syntactic and Semantic Features [J]. Journal of Computer Research and Development, 2016, 53(2): 284-302. (in Chinese)
甘丽新,万常选,刘德喜,等. 基于句法语义特征的中文实体关系抽取[J]. 计算机研究与发展, 2016, 53(2): 284-302.
- [4] CHOI S P, LEE S, JUNG H, et al. An intensive case study on kernel-based relation extraction[J]. Multimedia Tools & Applications, 2014, 71(2): 741-767.
- [5] ZENG D, LIU K, LAI S, et al. Relation Classification via Convolutional Deep Neural Network[C]// Proceedings of the 25th International Conference on Computational Linguistics. 2014: 2335-2344.
- [6] LI F L, KE J. Research Progress of Entity Relation Extraction Base on Deep Learning Framework [J]. Information Science, 2018, V36(3): 169-176. (in Chinese)
李枫林,柯佳. 基于深度学习框架的实体关系抽取研究进展[J]. 情报科学, 2018, V36(3): 169-176.
- [7] HOFFMANN R, ZHANG C, LING X, et al. Knowledge-based weak supervision for information extraction of overlapping relations[C]// Proceedings of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 2011: 541-550.
- [8] SURDEANU M, TIBSHIRANI J, NALLAPATI R, et al. Multi-instance multi-label learning for relation extraction[C]// Proceedings of Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012: 455-465.
- [9] BENJAMIN R, DIETRICH K. Combining Generative and Discriminative Model Scores for Distant Supervision[C]// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013: 24-29.
- [10] ZENG D, LIU K, CHEN Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks [C]// Proceedings of Conference on Empirical Methods in Natural Language Processing. 2015: 1753-1762.
- [11] JIANG X, WANG Q, LI P, et al. Relation Extraction with Multi-instance Multi-label Convolutional Neural Networks [C]// 26th International Conference on Computational Linguistics. 2016: 1471-1480.
- [12] LEI K, CHEN D, LI Y, et al. Cooperative Denoising for Distantly Supervised Relation Extraction [C]// Proceedings of the 27th International Conference on Computational Linguistics. 2018: 426-436.
- [13] BAHDANAU D, CHO K, BENGIO Y, et al. Neural Machine Translation by Jointly Learning to Align and Translate [J]. International conference on learning representations, 2015, 12(7): 366-381.
- [14] LIN Y, SHEN S, LIU Z, et al. Neural relation extraction with selective attention over instances [C]// Proceedings of Meeting of the Association for Computational Linguistics. 2016: 2124-2133.
- [15] JI G, LIU K, HE S, et al. Distant Supervision for Relation Extraction with Sentence-level Attention and Entity Descriptions [C]// Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. 2017: 3060-3066.
- [16] HE D, ZHANG H, HAO W, et al. A Customized Attention-Based Long Short-Term Memory Network for Distant Supervised Relation Extraction [J]. Neural Computation, 2017, 27(7): 1964-1985.
- [17] PENNINGTON J, SOCHER R, MANNING C. Glove: Global vectors for word representation [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014: 1532-1543.
- [18] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient Estimation of Word Representations in Vector Space [OL]. <https://arxiv.org/abs/1301.3781>.
- [19] CHUNG J, GULCEHRE C, CHO K, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling [J]. arXiv: 1412.3555, 2014.
- [20] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. Computer Science, 2012, 3(4): 212-223.
- [21] RIEDEL S, YAO L, MCCALLUM A, et al. Modeling relations and their mentions without labeled text [C]// European Conference on Machine Learning. 2010: 148-163.