计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2024,33(4):202−208 [doi: 10.15888/j.cnki.csa.009442]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

融合实体特征及多种类注意力机制的领域关系抽取模型①

王　稳, 刘远兴, 吴湘宁, 李文炽, 涂　雨, 张　锋, 方　恒, 蔡泽宇

(中国地质大学 (武汉) 计算机学院, 武汉 430078)
通信作者: 吴湘宁, E-mail: wxning@cug.edu.cn

摘　要: 基于远程监督的关系抽取方法可以明显地减少人工标注数据集的成本, 已经被广泛应用于领域知识图谱的构建任务中. 然而, 现有的远程监督关系抽取方法领域针对性不强, 同时也忽略了对领域实体特征信息的利用. 为了解决上述问题, 提出了一种融合实体特征和多种类注意力机制的关系抽取模型 PCNN-EFMA. 模型采用远程监督和多实例技术, 不再受限于人工标注. 同时, 为了减少远程监督中噪声的影响, 模型使用了句子注意力和包间注意力这两类注意力, 并在词嵌入层和句子注意力中融合实体特征信息, 增强了模型的特征选择能力. 实验表明, 该模型在领域数据集上的 *PR* 曲线更好, 并在 *P@N* 上的平均准确率优于 PCNN-ATT 模型.

关键词: 关系抽取; 知识图谱; 注意力机制; 实体特征

引用格式: 王稳,刘远兴,吴湘宁,李文炽,涂雨,张锋,方恒,蔡泽宇.融合实体特征及多种类注意力机制的领域关系抽取模型.计算机系统应用,2024,33(4):202–208. http://www.c-s-a.org.cn/1003-3254/9442.html

Domain Relationship Extraction Model Integrating Entity Feature and Multiple Types of Attention Mechanisms

WANG Wen, LIU Yuan-Xing, WU Xiang-Ning, LI Wen-Chi, TU Yu, ZHANG Feng, FANG Heng, CAI Ze-Yu

(School of Computer Science, China University of Geosciences, Wuhan 430078, China)

Abstract: The relationship extraction method based on remote supervision can cut the cost of labor-based annotated datasets and has been widely used in the construction of the domain knowledge graph. However, the existing remote supervised relationship extraction methods are not domain-specific and also neglect the utilization of domain entity feature information. To solve the above problems, this study proposes a relationship extraction model PCNN-EFMA that integrates entity features and multiple types of attention mechanisms. The model adopts remote supervision and multi-instance technology, no longer limited by labor-based annotation. At the same time, to reduce the impact of noise in remote supervision, the model uses two types of attention: sentence attention and inter-packet attention. In addition, it integrates entity feature information in the word embedding layer and sentence attention, enhancing the model's feature selection ability. Experiments show that the *PR* curve of this model is better on the domain dataset, and its average accuracy on *P@N* is better than that of the PCNN-ATT model.

Key words: relationship extraction; knowledge graph; attention mechanism; entity feature

1　引言

近年来, 与领域有关的自然语言文本数据快速增长. 为充分利用和挖掘这些自然语言文本数据中蕴含的知识, 需要对领域的文本数据进行知识抽取. 而关系

① 基金项目: 国家自然科学基金 (U21A2013); 智能地学信息处理湖北省重点实验室开放基金 (KLIGIP-2018B14)
收稿时间: 2023-07-02; 修改时间: 2023-09-09, 2023-10-25; 采用时间: 2023-11-09; csa 在线出版时间: 2024-03-04
CNKI 网络首发时间: 2024-03-08

抽取是知识抽取中一个重要的子任务, 是指从自然语言文本中提取实体之间的关系 (已事先定义种类), 即抽取出“实体-关系-实体”三元组结构. 关系抽取是构建知识图谱的必要步骤, 被广泛应用在机器问答、智能检索等方面.

在目前主流的关系抽取方法中, 有监督的关系抽取方法更为精准, 但是过于依赖人工标注的数据集. 而基于远程监督的关系抽取方法能够依靠远程知识库自动获得大量高质量的标注数据, 大大减少了人工标注的成本, 成为目前研究的热点.

然而, 基于远程监督的关系抽取方法在对齐远程知识库时, 往往会出现标记错误的问题, 导致模型的效果较差. 为了解决此类噪声对模型的影响, Zeng 等人[1]提出了多实例学习的关系抽取模型——PCNN (piecewise convolutional neural network) 模型, 通过多实例学习来抑制噪声对模型的影响. 但是以包为单位的多实例学习只选择包中概率最大的实例进行训练, 无法利用包中其他句子中的有效信息. 为了解决该问题, 基于句子注意力机制的 PCNN 模型被提出, 该模型在包中的所有实例上构建句子级别注意力, 为不同的句子赋予不同的权重, 这样就能综合利用包内所有句子的有效信息. 但是, 该方法忽略了句子中的头实体、尾实体这类词的特征, 在分配包内句子的注意力权重时, 没有考虑到头实体、尾实体对注意力权重的影响. 同时包内的句子注意力忽略了噪声包对模型的影响, 导致模型的效果并不理想.

基于上述研究现状, 本文结合实体特征及多种类注意力机制, 提出 PCNN-EFMA (piecewise convolutional neural network based on entity feature and multiple types of attention) 关系抽取模型. 模型为了解决包内句子注意力分配不合理的问题, 参照 TransE[2]算法的思想, 用头实体e_1和尾实体e_2可计算出实体之间的关系r, 即: $r \approx e_1 - e_2$. 因此, 对于同一关系, 理论上不同实体对计算的结果是相似的. 根据这一思想, 本文提出了一种融合领域实体信息的包内注意力计算方法, 该方法以头、尾实体为依据来计算包中不同句子的注意力权重, 降低了噪声对模型精度的影响, 提升了模型在垂直领域的适用性. 与此同时, 句子中的头、尾命名实体中也含有丰富的语义信息[3], 对于特定领域, 实体可能包含更多的信息. 所以, 在词嵌入层, 模型融合了头、尾实体向量来增强输入的语义信息, 使得模型能够充分利用所有的信息. 同时使用包间注意力机制来为不同的包分配权重, 避免噪声包对模型的影响.

2 相关研究

当前主流的关系抽取方法是基于深度学习的关系抽取算法, 可以分为两类: 有监督的关系抽取和远程监督的关系抽取. 有监督的关系抽取在训练时需要使用大量人工标注的数据集, 需要一定的人力成本. 而远程监督的关系抽取就是为了减少标注数据集的成本, 仅需要少量的标注数据集, 使用已有的知识库对语料进行自动标注, 并最终生成训练数据.

2009 年, Mintz 等人[4] 提出了基于远程监督的关系抽取方法, 通过与已有的远程知识库进行自然语言对齐, 实现自动标注. 远程监督的思想基于“如果两个实体在已知的知识库中存在某种关系, 那么所有提到这两个实体的句子都会以某种方式表达这种关系. ”这一假设, 而实际上在很多包含了两个实体的句子中其实并未包含那种关系, 从而导致远程监督得到的数据集中包含大量的噪声.

针对上述问题, Riedel 等人[5]在此基础上进行了改进, 提出了 EALO (expressed-at-least-once) 假设, 即“如果两个实体参与了一个关系, 那么至少有一个提到这两个实体的句子可能会表达这种关系”. 为了降低远程监督中噪声的影响, Riedel 等人[5]首次提出了多实例学习的思想, 将包含同一对实体的样本当作一个包, 模型只选择包中使得关系概率最大的一个实例作为实体对的表示. 基于这种思想, Zeng 等人[1]提出了基于远程监督的 PCNN 模型, 该模型以包为单位, 通过多实例学习来抑制噪声对模型的影响.

然而, 基于多实例学习的 PCNN 模型会丢失包中其他句子的关系, 导致特征遗漏. 因此, Lin 等人[6]将注意力机制引入 PCNN 模型中, 提出了用句子级别的注意力机制代替多实例学习. 该方法能够从包中的多个句子中学习特征, 将包中能够表示某种关系的句子被赋予较高的权重, 不能表示该关系的句子则被赋予较低的权重, 这样多个句子的特征就能够被学习到, 避免了特征遗漏的问题. 然而, 该方法却未考虑包样本中可能包含多种关系, 以及可能带有噪声的情况.

为了解决上述问题, Feng 等人[7]将强化学习运用于关系抽取领域, 提出了一种基于强化学习的关系抽取模型. 该模型分为样本选择器和关系分类器两部分,

其中在样本选择器中加入了强化学习, 对噪声实例和包进行移除, 有效地过滤远程监督中的噪声问题. 但是由于很多具有领域特征的数据被直接丢弃, 因而不适用于面向垂直领域的知识抽取.

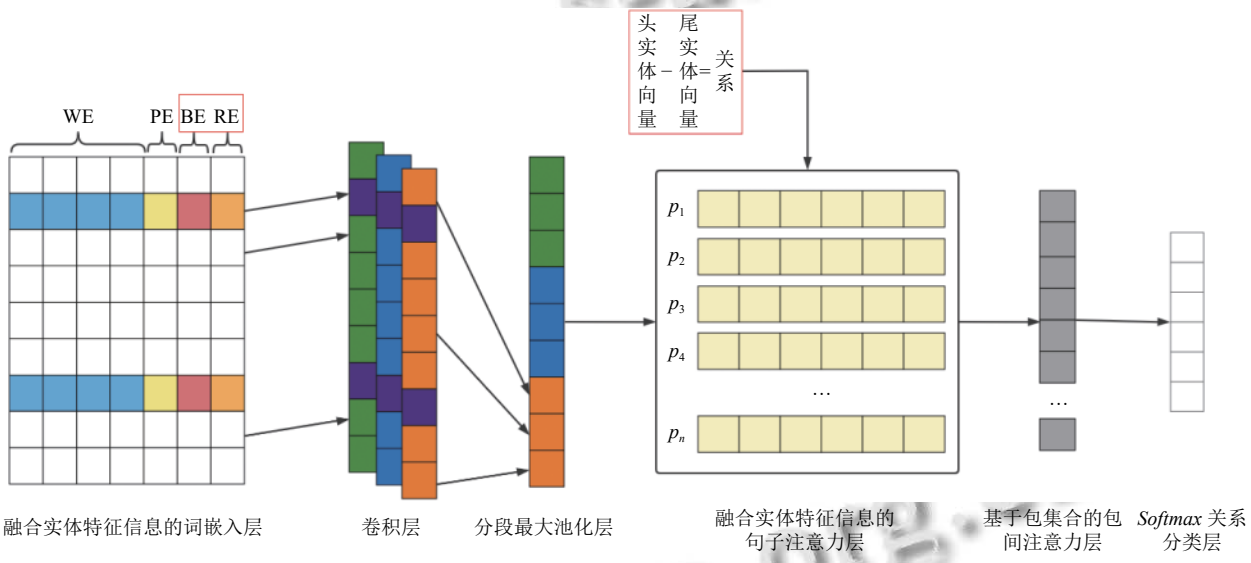
2019 年, Ye 等人[8] 提出了组合了包内注意力和包间注意力的多实例关系抽取模型, 先计算实体关系的权重, 再将拥有相同关系的包视为一个包集合, 然后基于包集合计算包间的注意力, 这样做不但可以减少噪声包的影响, 同时又能保证特征不丢失. 但是该方法同样存在没有利用领域实体信息的缺陷.

可见, 关系抽取算法仍存在一些需要解决的问题. 首先, 模型大都使用单一的注意力机制, 忽略了不同种类噪声包对模型效果的影响; 其次, 现有的关系抽取方法没有考虑到领域实体特征信息对关系分类的影响. 针对上述问题, 本文提出了一种既利用了实体特征, 又使用了多种类注意力机制 (句子注意力和包间注意力) 的关系抽取模型, 以提高针对特定领域关系抽取的性能.

3 PCNN-EFMA 模型

3.1 模型的结构

PCNN-EFMA 模型 (如图 1) 共分为 6 层: 融合实体特征信息的词嵌入层、卷积层、分段最大池化层、融合实体特征信息的句子注意力层、基于包集合的包间注意力层、*Softmax* 关系分类层.



WE: word embedding, PE: position embedding, BE: BIO embedding, RE: relation embedding

图 1 PCNN-EFMA 模型体系结构图

在融合实体特征信息的词嵌入层, 使用 Word2Vec 算法对输入的数据样本进行向量映射得到词向量, 然后计算句子中每个单词与头、尾实体的相对位置信息得到位置向量, 同时为了进一步利用头、尾实体在语句中的特征信息, 将头、尾实体向量与词向量及位置向量进行连接, 得到融合了词信息、位置信息和实体特征信息的特征向量.

在卷积层, 使用不同卷积核, 对输入的特征向量进行卷积运算, 以提取语句的不同特征.

在分段最大池化层, 按照头实体和尾实体的位置分为 3 段分别进行最大池化来提取特征值, 捕捉词语之间的语义关系.

在融合实体特征信息的句子注意力层, 参照 TransE 算法的思想来处理头实体和尾实体, 并以头、尾实体为依据来分配包中不同句子的注意力权重.

在包间注意力层, 将具有相同关系的包看作一个包集合来计算包间注意力, 从而克服噪声包对模型的影响.

Softmax 关系分类层用于对关系进行分类.

3.2 融合实体特征信息的词嵌入层

在命名实体识别任务中, 可以采用基于 Transformer 的 BERT 模型训练字向量输入到模型中进行特征提取. 基于字向量的 BERT 模型不适合用于关系抽取任务, 这里使用经典的 Word2Vec 模型作为词嵌入模型, 将

文本转换为数值向量. 除了使用词向量之外, 在关系抽取任务中, 距离头、尾实体不同远近的词往往会影响关系的预测结果. 通常, 离头、尾实体越近的词, 对关系的预测具有更高影响力, 因此, 在实现词嵌入时, 加入了词的位置信息, 构成位置向量.

同时, 根据 Ye 等人[3]的研究结果, 在关系抽取任务中, 头实体和尾实体是整个句子中最关键的两个词, 从整个句子的语义表达来讲, 这些关键的实体词包含着比其他词汇更为丰富的语义信息. 因此, 在词嵌入层, 在词向量、位置向量之后, 追加了头实体向量和尾实体向量, 使特征向量的语义更加丰富, 有利于模型挖掘更多信息.

(1) 词向量

Word2Vec 是基于分布式的静态词嵌入模型, 该模型是在大量无监督的语料上训练好词向量, 然后再将句子中的每一个词都映射到词表对应的向量. Word2Vec 有两种训练模式: CBOW 和 skip-gram. 其中, CBOW 模式是根据上下文去预测中心词, 而 skip-gram 模式则是根据中心词去预测上下文. 本文选用 skip-gram 的方式训练词向量. 假设词向量的长度为d_w, 则经过该方式训练词向量之后可以得到词向量矩阵$word \in R^{d_w \times |V|}$, 其中$V$表示单词的集合. 由句子 S 就可以得到表示句子的词向量集合$\{w_1, w_2, \cdots, w_n\}$, w_i表示句子中的第i个词语的向量表示, $w_i \in R^{d_w}$, n表示句子中词语的个数.

(2) 位置向量

位置向量最早由 Zeng 等人[9]提出并应用在关系抽取任务中, 取得了较好的效果. 位置向量记录每个词与头、尾实体词之间的相对距离, 有效地解决了神经网络无法学习到位置信息的缺点. 句子中的第i个词语, 与头实体和尾实体之间的相对距离可以用d_{i1}, d_{i2}进行表示, 其中 $d_{i1}, d_{i2} \in \{1, 2, 3, \cdots, n\}$. 如图 2 所示, 单词 CEO 与实体 Steve Jobs 和实体 Apple 之间的相对距离分别是 5 和−2. 然后再初始化两个位置向量矩阵PE_1, PE_2, 再把单词到头、尾实体的相对距离映射成低维度的位置向量, 得到pe_{1i}, pe_{2i}, $i \in \{1, 2, 3, \cdots, n\}$, 其中 $pe_{1i}, pe_{2i} \in R^{d_p}$.

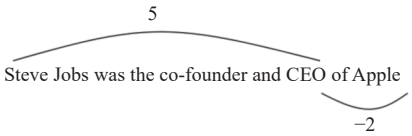


图 2 相对距离示例图

然后再将输入的句子$S = \{w_1, w_2, \cdots, w_n\}$中每个词的词向量和位置向量进行连接, 得到融合了位置向量的句子表示$SP = \{v_1, v_2, \cdots, v_n\}$, $SP \in R^{n \times d_x}$, 其中$v_i = [w_i; pe_{1i}; pe_{2i}]$, $d_x = d_w + 2 \times d_p$.

(3) 命名实体向量

在数据的预处理阶段, 会使用 BIO 的标记方式给句子中的每个词都进行标注. 对于给定的句子$S = \{w_1, w_2, \cdots, w_n\}$, 可以将命名实体标签转换成对应的向量表示$BE = \{bio_1, bio_2, \cdots, bio_n\}$.

(4) 头、尾实体向量

头、尾实体在关系抽取中具有非常重要的作用. 所以在词嵌入阶段, 将头、尾实体所表示的特征加入到输入层以强化头尾实体的作用. 根据 TransE 的思想, 采用头实体向量w_{e1}和尾实体向量w_{e2}两者的差值作为实体特征信息向量 re 拼接到输入向量中, 这里的 $re = w_{e1} - w_{e2}$.

然后再将融合了位置向量的句子表示$SP = \{v_1, v_2, \cdots, v_n\}$与命名实体向量$BE$和头、尾实体向量 re 进行拼接, 得到词嵌入层最终的输出向量$X = \{x_1, x_2, \cdots, x_n\}$, $X \in R^{n \times d_x}$, 其中$x_i = [w_i; pe_{1i}; pe_{2i}; bio_i; re]$, n 表示文本句子的长度, $d_x = 2 \times d_w + 2 \times d_p + d_{bio}$, 表示模型的输入向量维度, d_w, d_p, d_{bio}分别表示词向量维度、位置向量维度和命名实体向量维度. 向量X作为词嵌入层的输出, 输入到卷积层进行特征提取, 然后采用分段最大池化的方式得到句子的最后的特征向量P, 对于每一个包, 包内所有句子的特征可以表示为$B = \{P_1, P_2, \cdots, P_n\}$, n表示包内句子的数量, 再将B输入到融合实体特征信息的句子注意力层进行权重分配.

3.3 融合实体特征信息的句子注意力层

传统的以包为单位的多实例学习中, 只选择包中置信度最高的一个实例作为样本进行训练, 丢失了包中其他样本可能包含的有用信息. 所以在分段最大池化层之后, 一般会使用注意力机制来为包中不同的句子分配不一样的权重, 这样就能充分利用包中所有句子的有效信息. 然而现有的句子注意力机制在分配各个句子的权重时, 带有噪声的样本会影响权重的准确性, 难以正确选择包内不同句子的特征, 从而导致实体关系抽取的准确性下降.

针对上述问题, 在模型中使用了融合实体特征信息的句子注意力机制, 该方法参照 TransE 算法的思想,

实体关系标签r由头实体e_1和尾实体e_2的差值计算得出, 即$r \approx e_1 - e_2$. 将头实体、尾实体进行双线性变化, 再采用缩放注意力方法计算包中各个句子的权重, 最后对包中每个句子的特征加权求和后得到包的特征向量.

头实体向量v_{e1}和尾实体向量v_{e2}通过权重矩阵W^e实现双线性变化的公式如下, 这里b为偏置.

$$r = v_{e1} W^e v_{e2} + b \tag{1}$$

随后, 采用缩放注意力方法计算包中各个句子的权重. 实体包 B 经过分段最大池化层之后得到句子的特征表示$B = \{P_1, P_2, \cdots, P_n\}$, 对于包中的第$i$个句子特征$P_i$的权重$\alpha_i$可以用式 (2) 计算:

$$\alpha_i = Softmax\left(\frac{P_i^{\mathrm{T}} r}{\sqrt{d_k}}\right) \tag{2}$$

将包中每个句子的特征P_i加权求和, 得到包的特征向量P:

$$P = \sum\nolimits_i \alpha_i P_i \tag{3}$$

接着将包特征向量P进行线性变化, 获得包对每一类关系的置信度, 然后计算交叉熵损失, 再通过神经网络的反向传播对网络参数进行更新.

3.4 基于包集合的包间注意力层

句子注意力层能够充分利用包内所有句子的有效信息并抑制句子关系标记对模型的影响, 但是却无法解决实体关系标注错误的包, 即噪声包对模型的影响. 为了解决这个问题, 模型将拥有相同实体关系的包看作一个包集合, 在一组包集合内, 通过相似度计算包间注意力[8]. 具体计算公式如下:

$$G_k = \sum\nolimits_{i=1}^{n} \beta_{ik} s_k^i \tag{4}$$

$$\beta_{ik} = \frac{\exp(\mu_{ik})}{\sum_i \exp(\mu_{ik})} \tag{5}$$

其中, G_k表示关系标签为第 k 类的包集合, s_k^i表示第i个包B_i对于第k个类别的表示集合, β_{ik}表示第k类关系对于s_k^i的注意力权重, n表示集合中包的数量.

接着通过含有同一关系的包集合, 计算两个包之间的相关性, 公式如下:

$$\mu_{ik} = \sum\nolimits_{j \neq i} s_k^i s_k^{j\mathrm{T}}, \ s_k^i, s_k^j \in G_k \tag{6}$$

其中, s_k^i, s_k^j分别表示第i个和第j个包表示, μ_{ik}表示这两个包对于第k类关系的相关性.

因此, 包集合G_k被分到第k类关系的分值的计算公式如下:

$$\delta_k = r_k G_k + b_k \tag{7}$$

其中, δ_k是G_k分类到r_k的得分, b_k为偏置. 最后, 通过 *Softmax* 计算包的分类概率:

$$p(k \mid G_k) = \frac{\exp(\delta_k)}{\sum_i^{|R|} \exp(\delta_i)} \tag{8}$$

基于式 (8) 进行交叉熵计算得到损失函数, 并对模型进行训练.

4 实验结果与分析

4.1 实验数据集

PCNN-EFMA 关系抽取模型实验使用远程监督的方法进行指挥控制领域关系抽取数据集的构建, 原始数据集源自项目所收集到的指挥控制领域数据集, 再使用 CN-DBpedia 作为远程数据集进行对齐, 最后经过人工处理后, 得到 23 608 个句子和 1 628 个关系包. 然后按照表 1 所示的比例划分了训练集与测试集.

表 1 数据集详情

数据集	句子数	包个数	关系类别数
训练集	20296	1367	5
测试集	3312	261	5

4.2 实验参数设置

在 PCNN-EFMA 关系抽取模型的实验中, 使用网格搜索法来进行参数调优, 在实验中, 根据现有的研究结果, 将滑动窗口的取值大小设置为{3,4,5}, 卷积核数目的取值设置为{100,200,230,250,300}, 批处理数目的取值设置为{10,50,100,150}, 学习率的取值设置为{0.001,0.01,0.1}, 同时为避免模型过拟合, 将神经元的随机失活概率设置为{0.3,0.5,0.7}. 模型最终使用的参数如表 2 所示.

PCNN-EFMA 模型训练分两个步骤进行, 先在融合了实体特征的句子注意力层上训练, 然后在基于包集合的包间注意力层上再次训练.

4.3 实验结果分析

使用 Lin 等人于 2016 年提出的同样基于多实例学习的 PCNN-ATT[6]模型作为基准模型, 该模型同样使用 PCNN 网络进行特征提取, 使用句子注意力来为包内所有的句子分配权重, 以充分利用包内所有句子

的有效信息. 与 PCNN-EFMA 模型相比, PCNN-ATT 模型在词嵌入层没有融合领域实体特征信息, 在包内句子注意力的分配上, 也没有利用领域实体特征信息进行注意力权重的计算, 同时也没有考虑到噪声包对模型的影响, 因此没有使用基于包集合的包间注意力机制.

表 2 模型的参数表

参数	取值
词向量的维度设置	60
位置向量的维度设置	5
命名实体向量的维度设置	60
最大句子长度设置	150
滑动窗口大小	3
卷积核数目	230
句子注意力批处理数目	50
包间注意力批处理数目	10
包集合中包的最大个数	5
失活概率	0.5
学习率	0.001

在实验过程中, 对加入了实体特征及句子注意力方法得到的模型, 以及使用了包间注意力方法后得到的模型分别进行实验评估, 以验证两种方法的有效性. 于是选择 PCNN-EA 模型进行辅助验证, 该模型使用了融合了实体特征信息的句子注意力机制, 但是没有使用包间注意力机制.

实验采用准确度 (Precision)、召回率 (Recall) 和 *PR* 曲线 (Precision-Recall curve), 以及 *P@N* 指标来评估模型. 其中, *PR* 曲线是取不同的阈值进行实验之后, 将不同阈值下的准确度和召回率连接成一条曲线. *P@N* 表示前 *N* 个结果准确度的平均值, 如式 (9) 所示, 其中, y_i 表示第 i 个样本的准确度, 若为 0 则表示该样本抽取的关系不准确, 与标注的关系不符, 若为 1 则表示该样本抽取的关系准确, 与标注的关系一致. 例如 *P@*100 表示前 100 个结果准确度的平均值.

$$P@N = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{9}$$

实验结果如图 3 所示. 对比 PCNN-ATT 和 PCNN-EA 模型的实验结果可以发现, 融合了实体特征信息的句子注意力机制的 PCNN-EA 模型在性能上明显优于使用了普通的句子注意力机制的 PCNN-ATT 模型. 这表明在句子注意力权重的分配上, 将实体特征信息作为计算注意力权重的依据, 可以为关系预测提供更多的有效特征, 同时也让模型在句子特征选择上更加合理, 验证了本文提出的在词嵌入层和句子注意力层嵌入实体特征, 以及将句子注意力与包间注意力机制相结合方法的有效性.

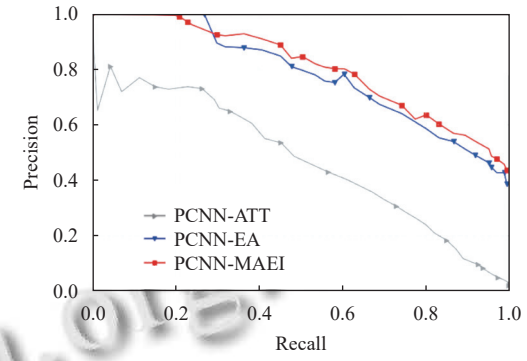


图 3 模型效果对比图

对比 PCNN-EA 和 PCNN-EFMA 模型的实验结果可知, 在相同准确率下, PCNN-EFMA 模型具有更好的召回率; 在相同的召回率下, PCNN-EFMA 模型具有更高的准确率. 实验结果说明, 基于包集合的包间注意力机制有利于发现包间潜在关系, 并降低噪声包对关系抽取准确性的影响.

对 one、two、all 示例的分析实验结果如表 3 所示. 由实验结果可知, PCNN-EFMA 模型在准确率和召回率上较基准模型 PCNN-ATT 有较大的提升, 模型的 *P@*100、*P@*200 和 *P@*300 指标均达到最佳. 该实验结果验证了本文提出的融合领域实体特征和多种类注意力机制的 PCNN-EFMA 关系抽取模型的有效性.

表 3 模型 *P@N* 评估表 (%)

示例	模型	*P@*100	*P@*200	*P@*300	平均值
One	PCNN-ATT	74.2	69.3	64.7	69.4
	PCNN-EA	81.5	73.6	69.4	74.8
	PCNN-EFMA	**84.1**	**75.4**	**71.0**	**76.8**
Two	PCNN-ATT	76.1	73.3	69.4	72.9
	PCNN-EA	82.6	80.5	74.3	79.1
	PCNN-EFMA	**86.6**	**78.4**	**74.9**	**80.0**
All	PCNN-ATT	82.5	78.7	75.6	78.9
	PCNN-EA	90.1	85.3	78.0	84.5
	PCNN-EFMA	**89.9**	**84.2**	**80.9**	**85.0**

5 结束语

本文结合实体特征以及多种类注意力方法, 提出了一种关系抽取模型 PCNN-EFMA. 该模型将领域实体信息融入词嵌入层, 提高了模型的领域适配性. 模型还使用句子注意力和包间两类注意力机制来减少远程

监督中噪声对模型的影响. 在句子注意力机制中, 模型使用了 TransE 算法构建特征向量的方法, 提出了融合实体特征信息的句子注意力机制, 使得模型在分配注意力权重时更加关注头、尾实体的特征, 解决了传统的注意力机制对句子特征选择不足的问题. 在包间注意力机制中, 模型将具有相同实体关系的包看作一个包集合, 通过对包分配包间注意力, 有效降低标注错误的包对抽取精确度的影响. 同时为了增强头实体和尾实体的特征, 在词嵌入层还将头实体和尾实体向量作为额外的特征拼接到 Word2Vec 得到的词向量上, 以丰富输入文本的语义信息, 让模型挖掘更多的潜在信息. 实验结果表明, PCNN-EFMA 模型在领域数据集上的 *PR* 曲线更好, 并在 *P@N* 上的平均准确率优于传统的 PCNN-ATT 模型.

参考文献

1 Zeng DJ, Liu K, Chen YB, *et al*. Distant supervision for relation extraction via piecewise convolutional neural networks. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: Association for Computational Linguistics, 2015. 1753–1762.

2 Bordes A, Usunier N, Garcia-Duran A, *et al*. Translating embeddings for modeling multi-relational data. Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc., 2013. 2787–2795.

3 Ye W, Li B, Xie R, *et al*. Exploiting entity BIO tag embeddings and multi-task learning for relation extraction with imbalanced data. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 1351–1360.

4 Mintz M, Bills S, Snow R, *et al*. Distant supervision for relation extraction without labeled data. 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. Suntec: Association for Computational Linguistics, 2009. 1003–1011.

5 Riedel S, Yao LM, McCallum A. Modeling relations and their mentions without labeled text. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Barcelona: Springer, 2010. 148–163.

6 Lin YK, Shen SQ, Liu ZY, *et al*. Neural relation extraction with selective attention over instances. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin: Association for Computational Linguistics, 2016. 2124–2133.

7 Feng J, Huang ML, Zhao L, *et al*. Reinforcement learning for relation classification from noisy data. Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans: AAAI, 2018. 5779–5786.

8 Ye ZX, Ling ZH. Distant supervision relation extraction with intra-bag and inter-bag attentions. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 2810–2819.

9 Zeng DJ, Liu K, Lai SW, *et al*. Relation classification via convolutional deep neural network. Proceedings of the 25th International Conference on Computational Linguistics. Dublin: Association for Computational Linguistics, 2014. 2335–2344.

(校对责编: 孙君艳)