

基于动态标签的关系抽取方法

薛露, 宋威*

(江南大学物联网工程学院, 江苏 无锡 214122)

(* 通信作者电子邮箱 songwei@jiangnan.edu.cn)

摘要:针对远程监督数据集的关系抽取研究方法存在着大量标签噪声的问题,提出了一种作用于分层注意力机制关系抽取模型的动态标签方法。首先,提出了一种根据关系类别相似性生成动态标签的概念。由于相同的关系标签包含相似的特征信息,计算特征信息的关系类别相似性有助于生成与特征信息相对应的动态标签。其次,利用动态标签方法的评分函数来评价远程监督标签是否为噪声,以决定是否需要生成新的标签代替远程监督标签,通过调整远程监督标签来抑制标签噪声对模型的影响。最后,根据动态标签来更新分层注意力机制以关注有效实例,重新学习每个有效实例的重要性,进一步抽取关键的关系特征信息。实验结果表明,相较于原始的分层注意力机制关系抽取模型,所提方法在 Micro 和 Macro 分数上分别有 1.3 个百分点和 1.9 个百分点的提升,实现了噪声标签的动态纠正,提升了模型的关系抽取能力。

关键词:关系抽取;远程监督;动态标签方法;评分函数;分层注意力机制

中图分类号:TP389.1 **文献标志码:**A

Relation extraction method based on dynamic label

XUE Lu, SONG Wei*

(School of Internet of Things Engineering, Jiangnan University, Wuxi Jiangsu 214122, China)

Abstract: Concerning the problem that the research methods of relation extraction for distant supervision datasets have a lot of label noise, a dynamic label method applied to the hierarchical attention mechanism relation extraction model was proposed. Firstly, a concept of generating dynamic label based on the similarity of relation categories was proposed. Since the same relation labels contain similar feature information, calculating the similarity of relation categories of feature information is helpful to generate the dynamic label corresponding to the feature information. Secondly, the scoring function of the dynamic label was used to evaluate whether the distant supervision label was noise and to determine whether a new label was needed to generate to replace the distant supervision label, and the influence of label noise on the model was suppressed by adjusting the distant supervision label. Finally, according to the dynamic label, the hierarchical attention mechanism was updated to focus on the effective instances, the importance of each effective instance was relearned and key relation feature information was further extracted. The experimental results indicate that, compared with the original hierarchical attention mechanism relation extraction model, the proposed method has the Micro and Macro scores increased by 1.3 percentage points and 1.9 percentage points respectively, realizes the dynamic correction of the noise label, and improves the relation extraction ability of the model.

Key words: relation extraction; distant supervision; dynamic label method; scoring function; hierarchical attention mechanism

0 引言

关系抽取(Relation Extraction, RE)旨在根据纯文本预测关系事实,由于抽取效果的好坏直接影响后续任务的准确率,这使得其在推荐系统、问答系统、智能搜索等自然语言应用领域占据重要地位^[1]。传统的监督 RE 模型缺乏标注的训练数据,手动标记训练数据既耗时且成本高。为此, Mintz 等^[2]提出了远程监督,通过将纯文本中的实体对与知识库(Knowledge Base, KB)带有关系标记的实体对直接对齐标注

自动生成训练数据。若 KB 中某个实体对之间无关系,则将待标注的句子标记为负实例(Negative Instance, NA)。例如,“Asmara”“Eritrea”在 KB 中存在 capital 的关系,则包含这两个实体的非结构化文本“I eventually fly from Nairobi to Dubai to Asmara, the capital of Eritrea.”可以作为模型训练的有效实例。但是,当“Asmara”与“Eritrea”同时存在一个句子中并非表达 capital 关系时,这种没有考虑实体对具体语义背景的标注方法产生了大量噪声。为了减轻错误的标签问题,Zeng

收稿日期:2019-11-18;修回日期:2020-01-08;录用日期:2020-01-10。

基金项目:国家自然科学基金资助项目(61673193);中央高校基本科研业务费专项资金资助项目(JUSRP51635B);中国博士后科学基金资助项目(2017M621625);江苏省自然科学基金资助项目(BK20181341)。

作者简介:薛露(1994—),女,重庆人,硕士研究生,主要研究方向:自然语言处理、关系抽取;宋威(1981—),男,湖北恩施人,副教授,博士,CCF 会员,主要研究方向:数据挖掘、人工智能。

等^[3]首先将基于卷积神经网络的深度学习模型运用于此项任务,与传统机器学习模型相比取得了显著的进步。

虽然研究者已经做出许多努力,而大多数现有方法在训练过程中使用远程监督标签,导致不能准确地获取关系特征信息,不能完全解决错误的标注问题。针对以上问题,本文提出了动态标签方法,可以通过某些实例与潜在正确标注实例之间的关系类别相似性评价远程监督标签的可靠性,在训练阶段动态纠正错误的远程监督(Distant Supervision, DS)标签,从而提高关系抽取的准确性。

1 相关工作

近年来学者们已经提出许多缓解远程监督数据集噪声的方法。Riedel等^[4]和Hoffmann等^[5]提出多实例学习(Multi-Instances Learning, MIL),如图1所示。该方法将训练数据集划分为多个实体对集合,将关系标签直接赋予集合,并提出以下假设,如果实体对之间存在某种关系,则集合中至少一个句子必须反映该关系,同时刻画一个实体对可能存在多种关系的情况。

Ernst Haefliger died on Saturday in Davos.
Ernst Haefliger was born in Davos on July.
Ernst Haefliger was born in Davos, Switzerland.



/location/location/Place of Death

图1 远程监督集合

Fig. 1 Distant supervision set

注意力机制从众多信息中选择出特定目标的关键信息。在MIL的基础之上,Zeng等^[6]引入注意力机制从集合中选择一个最有效的实例作为集合的特征,但此方法导致大多数有效实例被忽略,损失了大量有效信息。Lin等^[7]使用注意力机制将集合中所有实例按照重要程度综合表示为集合的特征,缓解大量有效信息被丢失的问题。Zhou等^[8]提出层次注意力机制,首先对集合的所有句子进行句子级别关注选择最相关的句子,然后采用单词级别的关注抽取区分关系类别的关键单词,构建句子特征表示,并聚合这些句子特征为集合特征。同样的,Qu等^[9]提出单词级别的注意力机制来区分句子中每个单词的重要性,从而增加了这些关键单词的注意力权重,并且将目标实体词向量的语义信息作为补充特征,进一步丰富集合特征;相似的,Ji等^[10]将实体描述信息作为额外的背景知识补充到模型中。李浩等^[11]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。该模型通过双向门控神经网络获取句子特征;其次,通过引入词语层、句子层、关系层的注意力机制,在学习不同关系之间联系的同时,减少错误标签。冯建周等^[12]设计了一种新的注意力模型,该模型构建排序的组合句子向量,使模型彻底地抛弃噪声句子,而不是让噪声句子以一个较低权重参与计算。Han等^[13]根据关系标签的层次结构,关注关系之间丰富的语义信息,在关系标签的每一层结构上引用注意力机制,从而抽取丰富的层次结构表示集合特征。这些模型的提出表明了注意力机制处理特征信息的有效性。为了进一步缓解噪声问题,Zeng等^[14]引入强化学习概念,通过奖惩机制区分句子是否是有效实例。Wu等^[15]采用对抗训练机制,使模型在训练过程中对噪声更加敏感,增强模型的鲁棒性。Liu等^[16]提出一种基于置信度的软标签方法,该方法利

用来自正确标记集合的特征,在训练过程中动态地纠正错误的标签。Ru等^[17]则从语义角度出发,使用Jaccard算法选择一个核心依赖短语来表示句子关系类别,以过滤远程监督标签噪声。以上的关系抽取模型,虽对关系抽取能力有一定提升,但是大多数现有方法在训练过程中使用远程监督标签,导致不能准确地获取关系特征信息。

通过分析MIL的集合得出:1)大量的集合标注了NA标签,但实际较多实例存在有效关系信息。2)集合中的所有实例或者大多数实例实际上表达的关系信息与关系标签不符合。3)较多集合只有一个实例,且关系标签可能错误。由于大多数现有方法在训练过程中使用的是硬标签,潜在的有效实例被忽略,仍然存在噪声问题影响模型关系抽取能力。

针对上述问题,本文提出了一种可以根据关系类别相似性产生动态标签(Similarity dynamic Label, SL)的方法。本文将提出的SL方法作用于分层注意力机制(Hierarchical ATTention, HATT)^[13],提出基于动态标签的分层注意力机制(HATT based on SL, SL-HATT)RE模型,以验证动态标签的调整对RE产生积极的影响。首先,根据HATT对关系标签(例如/people/person/nationality)的层次逐层抽取集合的特征表示。其次,动态标签方法计算集合特征之间的关系类别相似性,并且结合远程监督标签的可信度,共同评价集合的关系标签是否标注正确,若标注错误,则根据评分函数产生新的动态标签,代替原来的远程监督标签作为训练时的标签,并且动态标签在训练过程中不断调整,以解决集合的噪声标签问题。最后,在训练过程中根据动态标签的调整,更新HATT关注集合的有效实例,以抽取更丰富的集合关键词特征。

SL方法与HATT的结合的优势在于两个方面:一方面,SL方法首次提出通过关系类别相似性调整集合的标签。另一方面,HATT根据新的标签,更新关注集合中有效实例,使得有效实例增多,以便于抽取区分关系类别的关键词特征信息。并且,HATT注意力机制主要对关系标签之间的层次关联建模,抽取了关系之间的相关性信息,使得关系标签之间不以独立的方式存在。二者共同作用于关系标签的层面,抽取关系之间丰富关联性的同时,根据关系之间的相似性调整错误标签,缓解集合的错误标签对关系抽取整个过程带来的影响。总的来说,从两个层面共同作用缓解远程监督的噪声问题。特别的,在多实例学习方法的基础之上,正确标注的标签远多于错误标签,动态标签思想得以实现^[16]。

2 SL-HATT模型

本文提出了一种根据关系类别相似性产生动态标签的方法。在训练过程中代替远程监督标签,减弱错误的远程监督标签对关系抽取能力的影响。为了验证本文提出的方法,将SL方法作用于HATT,提出SL-HATT模型,通过同时对关系标签进行动态调整来学习更充足的集合关系特征信息。图2是SL-HATT网络结构框架,由以下两部分组成:1)HATT关系抽取模型;2)SL方法。

2.1 句子编码器以及HATT

早期的RE方法主要依赖自然语言处理(Natural Language Processing, NLP)工具提取文本特征。近年来,神经网络因具有更准确地捕获文本特征的能力而被广泛用于RE。

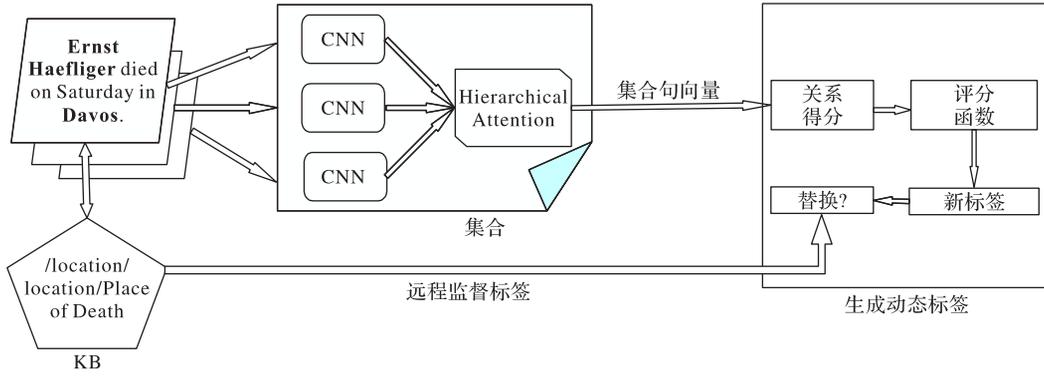


图2 SL-HATT网络结构框架

Fig. 2 SL-HATT network structural framework

在 NLP 任务中,通常以词向量的形式作为神经网络的输入。词向量是词的分布式表示,具有维度低、语义信息丰富的优点。本文采用 Skip-Gram 方法通过大型语料库训练得到的 50 维的词向量 w_e 作为输入, $w_e \in E^{n \times KW}$, E 为词向量矩阵, n 为词典长度, KW 为词向量维度,其中意思相近的词将被映射到向量空间中相近的位置,保证模型输入带有单词词义信息。同时,通过加入每个单词与实体在句子中的相对位置信息,可以对句子中不同的单词与实体之间的相对位置进行分布式表示^[3],提升关系抽取的效果。位置向量表示为 w_p ,每个词的向量由 w_e 和 w_p 组成,定义为 w 。其中,对于长度为 m 的句子实例,有式(1):

$$W = [w_1, w_2, \dots, w_m] \quad (1)$$

为了便于与其他关系抽取模型比较,本文选择卷积神经网络(Convolutional Neural Network, CNN)^[3]作为句子编码器,将实例矩阵 W 编码为句子的隐藏层特征矩阵 x ,如式(2)。然后,将最大化池化层作用于卷积层输出的 x ,以得到最终句子特征向量 h ,如式(3)所示。

$$x = CNN([w_1, w_2, \dots, w_m]) \quad (2)$$

$$h = \max(x) \quad (3)$$

由于关系标签在其层次结构上具有相关性,如/people/person/nationality 和 /people/person/ethnicity,高层次(例如 people)的关系更加通用,低层次(例如 nationality)的关系信息更特殊。Han 等^[13]提出,给定远程监督关系标签集 R ,通过关系层次生成关系层次集合 $\{R^1, R^2, \dots, R^k\}$,其中 k 为关系层数, R^k 为当前层次的关系集合。在关系层次结构的每一层上对集合进行注意力关注,以获得相应层次的注意力特征。对于一个包含 N 条实例的集合,计算针对远程监督标签的第 k 层关系集合的特征向量 r^k ,如式(4),其中“ \cdot ”表示点乘。 α_i^k 为集合中第 i 条句子在第 k 层关系集合的重要性权重,如式(5)。

$$r^k = \sum_{i=1}^N \alpha_i^k \cdot h_i \quad (4)$$

$$\alpha_i^k = \frac{\exp(q_k^T A h_i)}{\sum_j \exp(q_k^T A h_j)} \quad (5)$$

其中: A 为权重矩阵; q_k 为 R^k 中关系的查询向量。最后,将总共 k 层的特征如式(6)串联:

$$r = [r^1; r^2; \dots; r^k] \quad (6)$$

其中 r 表示集合的特征向量。

图3是 HATT 根据关系层次抽取特征的过程,对关系标签的层次逐层进行有效实例的关注,能够加强网络对文本信息的表达能力,抽取更加丰富的集合语义特征。

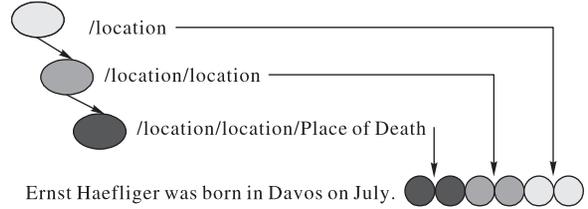


图3 HATT

Fig. 3 HATT

2.2 动态标签方法

远程监督数据集的噪声为以下几个方面:首先,包含大量 NA 标签的集合,但实际较多实例存在有效关系信息。其次, MIL 集合中大多数实例实际上表达的关系信息与关系标签不符合。最后,集合仅有一个实例且关系标签可能错误。对此,本文提出了动态标签方法,在训练期间根据集合特征信息之间的关系类别相似性决定是否生成新的标签代替原来的远程监督标签,以减少错误标注对关系抽取过程的影响。

对于 T 组集合,根据第 i 组集合特征向量 r_i 计算其关系得分向量 S_i ,定义如式(7)。通过关系权重矩阵 M 将集合特征映射到 R^o 维,其中, o 为关系类别数目, S_i 向量则代表集合 i 对应于每一类关系的得分。

$$S_i = \frac{\exp(Mr_i + b)}{\sum_{j=1}^T \exp(Mr_j + b)} \quad (7)$$

由评分函数决定集合 i 的动态标签热向量 L_i^d ,如式(8)~(9)所示:

$$I_i = \arg \max (sim(S_i) + (S_i + L_i)) \quad (8)$$

$$L_i^d = \text{one-hot}(I_i) \quad (9)$$

式(8)中: $\arg \max$ 函数计算索引值 I_i ,这里索引代表关系类别。式(9)生成动态标签的热向量(one-hot)表示 L_i^d ,以替代远程监督 L_i 。

由式(8),组成评分函数的两部分如下:第一部分,代表集合之间综合关系类别相似性;第二部分,代表集合远程监督标签的可信度。两部分组成一个综合表示向量,取最大值的索引 I_i ,该索引代表一个关系类别,生成新的动态标签。如果与远程监督标签相同,说明集合暂时不需要替换标签;反之,说明集合的标签需要被替换为 L_i^d 。两部分共同决定集合的动态标签,既考虑了远程监督标签的影响,又加入了关系类别相似性特征对远程监督标签的限制。训练过程中不断生成新的标签,直至收敛。

第一部分的关系类别相似性计算过程如下。分析集合关系得分向量,在训练过程中,拥有相同关系标签的集合,其关

系特征存在相似性。根据自注意力机制能够解决长距离依赖学习的思想,计算任意两两集合之间的关系类别相似性,并得到相似性权重值 $\alpha(S_i, S_j)$,根据 $\alpha(S_i, S_j)$ 大小决定与集合 i 相似的集合,选择集合的远程监督标签作为可能待产生的新标签,以改善集合的标签噪声问题。由此,存在一组集合的关系得分向量 S_i ,通过内积函数 $Y_s(S_i, S_j)$ 计算与其他集合之间相似性,并通过 Softmax 得到集合之间的相似性权重值 $\alpha(S_i, S_j)$,代表任意集合 i 与 j 之间的相似性,如式(10)所示。

$$\alpha(S_i, S_j) = \frac{\exp(Y_s(S_i, S_j))}{\sum_{j=1}^T \exp(Y_s(S_i, S_j))} \quad (10)$$

其中:权重 $\alpha(S_i, S_j)$ 越大,表明集合 i 与集合 j 之间的关系类别越相似; T 为当前集合的个数。

根据 $\alpha(S_i, S_j)$ 计算集合 i 属于某种关系类别的可能性。进一步,将其累加求和得到集合 i 的综合相似性表示向量 $sim(S_i)$,如式(11):

$$sim(S_i) = \sum_{j=1}^T \alpha(S_i, S_j) \cdot L_j \quad (11)$$

其中, L_j 为远程监督标签热向量,该部分首次将标签信息作为特征,计算集合 i 与其他集合的相似程度。由于数据集中较多集合的标签为 NA,可能导致生成动态标签时更偏向 NA 标签,所以屏蔽 $sim(S_i)$ 中 NA 标签以避免对生成动态标签过程的影响。由于 $sim(S_i)$ 包含集合 i 与其他集合的关系类别相似性特征,通过 $sim(S_i)$ 的最大值的索引作为可能待产生的新标签。

第二部分为远程监督标签的可信度,分为两个子部分(前者和后者)。前者 S_i 为集合 i 的关系得分向量;后者 L_i 为远程监督标签热向量。两者求和判断前者与后者的最大值索引是否相同,以判断关系类别得分与远程监督标签代表的关系类别是否相同。若 S_i 与远程监督标签表示关系相同,则信任远程监督标签,此部分将对关系标签不作调整;若不同,则削弱远程监督标签的影响,并增加关系得分 S_i 代表的关系标签可能性,对待产生的动态标签起促进作用。

图4为标签调整过程,其中: S 代表集合特征向量,圆圈表示关系类别,颜色深浅表示属于某种关系的可能性, L 代表远程监督关系标签。以集合 S_4 为例,当根据类别相似性语义生成一个新标签 L_2^d ,且关系得分 S_4 与远程监督标签的关系类别不同时,集合 S_4 的远程监督标签 L_4 被替换为 L_2^d ,说明集合 S_4 远程监督标签可能错误,可能与集合 S_1 和 S_3 更相似。替换完成后,以 L_2^d 训练模型参数。

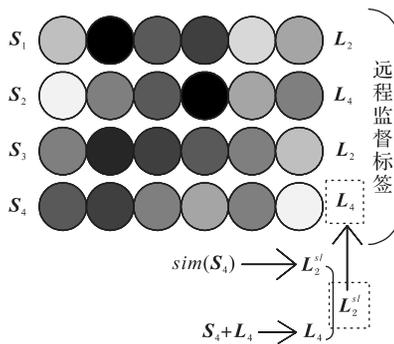


图4 生成动态标签的过程

Fig. 4 Process of generating dynamic label

由式(8)的两部分综合表示共同决定是否产生新的动态标签。新生成的动态标签作为训练时的标签,并且动态标签可能在训练过程中进行不断的调整,以更新 HATT 关注集

中有效实例。综上,不断进行的调整使得所有的集合都拥有更多的有效实例,以达到为远程监督数据集降噪的目的。

下面针对 SL-HATT 模型进行详细介绍:

1) 将集合文本单词以 Skip-Gram 词向量和位置向量表示为 w , 将句子矩阵表示为 W 。

2) 添加 MIL 层, 将训练数据集划分为多个实体对集合, 将关系标签直接赋予集合。

3) 添加 CNN 层和最大池化层, 利用式(2)~(3) 得到集合中的句子特征矩阵 x 和句子特征向量 h 。

4) 引入 HATT 层, 利用式(4)~(6) 将集合中的句子以标签层次关联的注意力机制, 抽取到集合中所有句子的综合关系类别特征、集合特征向量 r 。

5) 在少许训练次数后, 添加 SL 方法层, 利用式(7)~(11) 学习集合之间的关系类别相似性特征, 以及判断集合特征与远程监督标签特征是否相同, 决定是否需要生成新的标签热向量 L^d 。

6) 定义模型损失函数, 通过更新成参数矩阵, 达到优化模型的效果。

2.3 模型训练

本文模型训练的优化算法采用随机梯度下降(Stochastic Gradient Descent, SGD)算法, 损失函数为交叉熵代价函数。同时为了避免过拟合问题, 加入了 L_2 正则化。给定实体对集合, 以及相应的远程监督关系, 将损失函数定义如下,

$$loss = -\frac{1}{T} \sum_{i=1}^T \ln P(L_i^d | r_i; \theta) + \lambda \|\theta\|^2 \quad (12)$$

其中: T 集合数目; L_i^d 是动态标签; r_i 是集合 i 的特征向量; λ 是正则化参数; θ 为关系抽取模型中的训练参数。

3 实验与结果分析

3.1 数据集及参数设置

本文采用由远程监督方法生成的纽约时报(New York Times, NYT)数据集。其中,关系标签有 53 种,包含无用关系 NA。训练集有 570 088 条句子、293 162 个实体对集合,测试集有 172 448 条句子、96 678 个实体对集合。为了保证实验的准确性,测试集由人工手动标注。在训练和测试集中,将超过 120 个单词的句子截断为 120 个单词^[13]。

本文使用交叉验证来确定模型的参数,由于 SL 方法与基线模型相同的实体对集合,并且应用文献[13]的 HATT 网络结构。为了公正地比较 SL-HATT 模型与基线模型,本文采用文献[13]实验的大部分参数,与文献[7]使用相同的 CNN 编码器及参数、预训练的词向量,如表 1 所示。由于 SL 方法对与 HATT 结合的时间较敏感,实验过程中单独测试其关键参数全局步数(global_step)。

3.2 实验对比模型及评估策略

本文分别采用近年来几种经典的关系抽取模型与 SL-HATT 模型进行比较,相应的模型如下。

1) PCNN+ATT (Piecewise CNN ATTention)^[6] 是使用分段卷积神经网络结合普通注意力机制的关系抽取模型。

2) PCNN-ONE+Soft 是基于软标签方法同时只关注集合中的一个句子的注意力模型^[16]。

3) PCNN-ATT+Soft 是基于软标签的注意力模型^[16],该模型使用分段卷积作为句子编码器,结合注意力机制,首次提出软标签方法,提出相对简单的评分函数调整关系标签。

4) RL (Reinforcement Learning) 是一种利用强化学习在远程监督数据中进行句子关系抽取的模型^[14]。在实验中,都采

用了相同的常规参数以统一实验变量来获得更准确的实验结果。

5) CNN+HATT 是基于标签关系分层注意力机制的模型^[13], 该方法是基于以层次划分关系标签的注意力机制, 相较于大多数现有方法中每种关系都是彼此独立的, 该方法的关系层次结构很好地揭示了关系之间的丰富关联。

表 1 实验参数

Tab. 1 Experimental parameters

名称	变量名	值
批处理数量	<i>Batch_size</i>	160
学习率	<i>Learning_rate</i>	0.2
隐含层参数维度	<i>Hidden_size</i>	230
词向量维度	<i>KW</i>	50
位置向量维度	<i>KP</i>	5
丢失率	<i>Dropout</i>	0.5
卷积窗口大小	<i>L</i>	3

本文采用准确率(Precision, P)、微观平均 F1 分数(Micro)和宏观平均 F1 分数(Macro)对实验结果进行评估。Micro 是一种对数据集中的每一个实例不分类别进行统计, 从而建立全局混淆矩阵的评价指标。Macro 则是先对每一个类统计指标值, 然后再对所有类求算术平均值。

3.3 结果分析

3.3.1 SL 方法 *global_step* 讨论

本节中, 对模型中一项重要参数 *global_step* 做了中间实验来讨论其取值。由于 SL-HATT 模型动态产生新的标签具有一定的随机性。为了减少随机性, 且使模型拥有基本的关系抽取初始权重, 本文设置实验, 在开始训练一定次数的 *global_step* 后, SL 方法开始动态地调整关系标签。

如图 5 所示, 由于 *global_step* 次数设置太大会导致噪声标签在训练过程的累计影响太大, 本文取 *global_step* = [1 000, 3 000, 5 000, ..., 10 000], 根据实验得出, *step_5000* 时模型精确率-召回率(Precision-Recall, P-R)曲线表现较好, 故本文设置在第 5 000 次 *global_step* 时, 引入 SL 方法较适宜。由于第 10 000 次 *global_step* 时模型抽取能力有所下降, 文本不再继续对后面的 *global_step* 进行讨论。

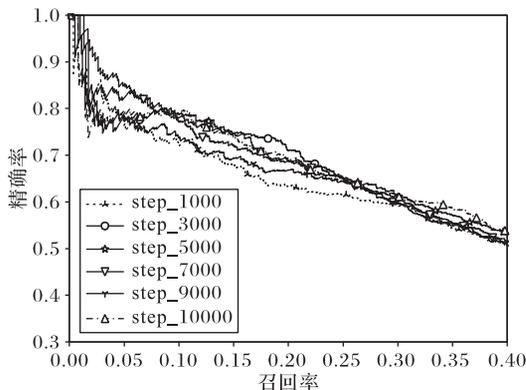


图 5 不同 *global_step* 次数比较

Fig. 5 Comparison of different *global_step* times

3.3.2 实验效果及模型性能分析

本节中, 主要对本文模型的性能进行详细的分析比较。分别将本文提出的 SL-HATT 模型与其他关系抽取模型比较, 其 P-R 曲线如图 6 所示。由图 6 可以看出, 本文提出的 SL 方法, 使得 SL-HATT 模型整体性能在 P-R 曲线中有着较为明显的优势。这反映出本文模型对远程监督数据集中噪声标签能

有效动态纠正, 提取到更加丰富有用的特征信息, 从而提升模型的整体效果。

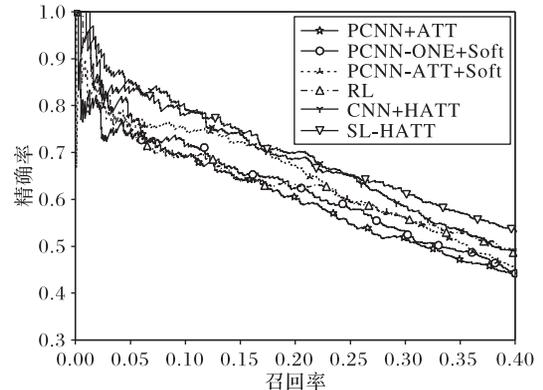


图 6 不同模型的 P-R 曲线

Fig. 6 P-R curves of different models

如图 6 所示, 本文提出的 SL-HATT 模型在 CNN+HATT 模型基础之上加入了 SL 方法, 相较于 CNN+HATT 模型关系抽取能力更强, 表明基于关系类别相似性的动态标签方法是可行且有效的, 对远程监督标签动态纠正实现了标签噪声的有效抑制, 加强了关系抽取能力。

特别地, 将本文提出的 SL-HATT 模型与 PCNN-ATT+Soft 模型进行比较。前者根据评分函数决定集合特征之间的相似性, 选择与之相似性最高集合的远程监督标签。后者仅根据集合关系特征向量与集合远程监督标签的置信度决定是否更新标签, 缺少集合之间的相似性计算, 本文认为根据集合之间的关系类别相似性获取的特征更准确, 便于关系标签调整。特别的, 由于 SL 方法与 HATT 的组合, 本文模型即使使用简单的句子编码器, 依然取得了较好的关系抽取效果, 表明关系标签拥有丰富的特征信息, 有助于集合特征抽取。

由于更加关注那些排名靠前的结果的性能, 研究召回率为 0.1、0.2、0.3 及其平均值时不同模型的精确率, 结果如表 2 所示。由表 2 可以看出: 一般的模型, 在提高召回率的同时, 会伴随着精确率的急剧下降, 而 SL 方法则同时兼顾了精确率与召回率, 表明本文提出的方法对远程监督数据集中的噪声处理使得模型的鲁棒性更强。

表 2 不同召回率下不同模型的精确率

Tab. 2 Precisions of different models under different recalls

召回率	精确率/%					
	PCNN+ ATT	PCNN- ONE+Soft	PCNN- ATT+Soft	RL	CNN+ HATT	SL-HATT
0.1	67.5	72.1	75.0	69.9	78.9	78.8
0.2	60.6	63.0	67.2	62.8	69.9	69.0
0.3	51.6	53.5	56.8	56.8	58.5	62.1
平均	59.9	62.8	66.3	63.1	69.1	70.0

由于数据集中各个类的分布不平衡, 本文使用 Macro 和 Micro 评价模型性能。由表 3 可知, 本文模型的 Micro 分数以及 Macro 分数相较基线模型(PCNN+ATT)有 5 个百分点和 3.3 个百分点的提升, 相较 CNN+HATT 模型有 1.3 个百分点和 1.9 个百分点的提升, 表明标签调整策略对模型整体性能发挥了积极的影响。特别的, 据统计, 在测试数据集中, 有 74 857 个集合只对应一个句子, 约占所有集合的 3/4^[7], 标签错误直接影响关系抽取的效果。因此, 对于仅含有一个实例的集合, SL 方法尤其重要。

表 3 微观平均分与宏观平均分 单位:%
Tab. 3 Mean values of Micro and Macro unit:%

模型	Micro	Macro	模型	Micro	Macro
PCNN+ATT	38.0	15.1	SL-HATT	43.0	18.4
CNN+HATT	41.7	16.5			

由于SL方法动态产生新的标签,可能存在实例的关系表达方式与常规关系模式有所不同,产生错误动态标签的情况。对此进行以下分析:首先,数据集中噪声标签远少于正确标注的标签。其次,在训练过程中根据集合的关系得分和远程监督标签生成动态标签,随着训练收敛过程,产生错误标签的可能性会越来越小。最后,本文在训练开始的5 000次 *global_step* 时开始引入SL方法,试图减少动态标签的随机性。根据实验结果来看,本文认为在训练过程中对关系事实进行较小的错误更正不会对整体性能产生太大影响。

综上,SL-HATT模型的优势在于两个方面:首先,SL方法改善了以往训练时使用硬标签的方式,根据集合关系类别相似性信息与远程监督标签相互对抗,决定是否调整关系标签,最终使得集合获得更合适的关系标签。其次,SL方法作用于HATT,不仅HATT抽取关系之间的丰富关联性,而且促进HATT动态地重新关注集合中的实例,使集合中包含更准确的关键词特征和有效实例特征。二者共同作用于关系标签的层面,共同加强关系抽取能力。

4 结语

本文针对远程监督数据集噪声较多的问题,提出了一种根据关系类别相似性产生动态标签的方法。将本文提出的动态标签方法作用于分层注意力机制,二者同时作用于关系标签的层面,不仅抽取关系之间丰富的关联性,而且根据关系之间的相似性动态调整关系标签,更新注意力机制关注有效实例,抽取更全面以及丰富的句子表征,明显加强了模型的关系抽取能力。实验表明,SL-HATT模型相较以往关系抽取模型性能更优,即使是简单的句子编码器也能达到较好的关系抽取效果。在以后的研究中,我们将继续在句子编码器方面进一步优化SL-HATT模型。

参考文献 (References)

- [1] 刘峤,李杨,段宏,等. 知识谱图构建技术综述[J]. 计算机研究与发展, 2016, 53(3):582-600. (LIU Q, LI Y, DUAN H, et al. Knowledge graph construction techniques [J]. Journal of Computer Research and Development, 2016, 53(3): 582-600.)
- [2] MINTZ M, BILLS S, SNOW R, et al. Distant supervision for relation extraction without labeled data [C]// 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 AF-NLP. Stroudsburg: ACL, 2009:1003-1011.
- [3] 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: Technical Papers. Stroudsburg: ACL, 2014: 2335-2344.
- [4] RIEDEL S, YAO L, MCCALLUM A. Modeling relations and their mentions without labeled text [C]// Proceedings of the 2010 Joint European Conference on Machine Learning and Knowledge Discovery in Databases, LNCS 6323. Berlin: Springer, 2010: 148-163.
- [5] HOFFMANN R, ZHANG C, LING X, et al. Knowledge-based weak supervision for information extraction of overlapping relations [C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies.

Stroudsburg: ACL, 2011: 541-550.

- [6] ZENG D, LIU K, CHEN Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks [C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2015: 1753-1762.
- [7] LIN Y, SHEN S, LIU Z, et al. Neural relation extraction with selective attention over instances [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2016: 2124-2133.
- [8] ZHOU P, XU J, QI Z, et al. Distant supervision for relation extraction with hierarchical selective attention [J]. Neural Networks, 2018, 108:240-247.
- [9] QU J, OUYANG D, HUA W, et al. Distant supervision for neural relation extraction integrated with word attention and property features [J]. Neural Networks, 2018, 100:59-69.
- [10] JI G, LIU K, HE S, et al. Distant supervision for relation extraction with sentence-level attention and entity descriptions [C]// Proceedings of the 31st AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2017: 3060-3066.
- [11] 李浩,刘永坚,解庆,等. 基于多层次注意力机制的远程监督关系抽取模型[J]. 计算机科学, 2019, 46(10): 252-257. (LI H, LIU Y J, XIE Q, et al. Distant supervision relation extraction model based on multi-level attention mechanism [J]. Computer Science, 2019, 46(10): 252-257.)
- [12] 冯建周,宋沙沙,王元卓,等. 基于改进注意力机制的实体关系抽取方法[J]. 电子学报, 2019, 47(8): 1692-1700. (FENG J Z, SONG S S, WANG Y Z, et al. Entity relation extraction based on improved attention mechanism [J]. Acta Electronica Sinica, 2019, 47(8): 1692-1700.)
- [13] HAN X, YU P, LIU Z, et al. Hierarchical relation extraction with coarse-to-fine grained attention [C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2018: 2236-2245.
- [14] ZENG X, HE S, LIU K, et al. Large scaled relation extraction with reinforcement learning [C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2018: 33-41.
- [15] WU Y, BAMMAN D, RUSSELL S. Adversarial training for relation extraction [C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2017: 1778-1783.
- [16] LIU T, WANG K, CHANG B, 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. Stroudsburg: ACL, 2017: 1790-1795.
- [17] RU C, TANG J, LI S, et al. Using semantic similarity to reduce wrong labels in distant supervision for relation extraction [J]. Information Processing and Management, 2018, 54(4):593-608.

This work is partially supported by the National Natural Science Foundation of China (61673193), the Fundamental Research Fund for the Central Universities (JUSRP51635B), the China Postdoctoral Science Foundation (2017M621625), the Natural Science Foundation of Jiangsu Province (BK20181341).

XUE Lu, born in 1994, M. S. candidate. Her research interests include natural language processing, relation extraction.

SONG Wei, born in 1981, Ph. D., associate professor. His research interests include data mining, artificial intelligence.