

融合门控机制的远程监督关系抽取方法

李兴亚 陈钰枫[†] 徐金安 张玉洁

北京交通大学计算机与信息技术学院, 北京 100044

[†] 通讯作者, E-mail: chenylf@bjtu.edu.cn

摘要 在关系抽取研究领域中, 远程监督是获取大规模标注语料的一种流行方式。该方法虽然在一定程度上减少了模型对人工标注数据的依赖, 但是依赖于实体标注工具的性能, 且存在关系标签标注错误的问题, 容易引入大量噪声数据。已有的研究工作主要集中在句子级别的噪声过滤, 即先获取句子的整体特征后, 再对句子进行筛选。虽然这种方式取得了较好的成效, 但是忽视了词级别的噪声过滤, 且模型训练过程中标签的不可变性会扩大错误标签所带来的影响。针对上述问题, 提出一种融合门控机制的远程监督关系抽取方法, 在词级别上自动选择正相关特征, 过滤与关系标签无关的词级别噪声; 并在门控机制内引入软标签的思想, 弱化硬标签对噪声过滤的影响; 最后结合句子级别的噪声过滤, 最终提升模型的整体性能。在公开数据集上的实验结果表明, 相对于句子级别噪声过滤方法, 所提方法的性能取得了显著提高。

关键词 关系抽取; 远程监督; 门控机制; 卷积神经网络

中图分类号 TN914

Distant Supervision for Relation Extraction with Gate Mechanism

LI Xingya, CHEN Yufeng[†], XU Jinan, ZHANG Yujie

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044

[†]Corresponding Author, E-mail: chenylf@bjtu.edu.cn

Abstract Distant supervision reduces the dependence on the manual annotated data for relation extraction, but suffers from wrong labels, which would introduce lots of noisy instances. The existing research work mainly focuses on sentence-level noise filtering, that is, the sentence will be selected after obtaining the features of the sentence. However, these methods ignore noise words in sentences, and hard labels, which are immutable during training, would worsen wrong labelling problem. To solve the above problems, a piecewise convolutional neural network with gating mechanism is proposed, which would automatically filter positive correlation features at word-level. Moreover, the idea of soft-label is introduced to the gating mechanism, which will weaken the impact of hard labels on noise filtering. Combined with sentence-level noise filtering, the overall performance of the model is improved. The experimental results on the public dataset show that the proposed model has a significant improvement compared to the sentence-level noise filtering methods.

Key words relation extraction; distant supervision; gated convolutional networks

关系抽取任务旨在从自然语言文本中识别出两个实体间所存在的语义关系, 是信息抽取任务的重要组成部分。目前关系抽取任务面临的一大挑战是模型训练需要大规模的标注语料, 而人工标注语料过于费时费力。为此, Mintz 等^[1]于 2009 年提出了远程监督关系抽取方法。远程监督使用知识库中的关系对训练语料进行启发式的标注。例如, 给定一个知识库中的元组, 也称为关系事实, “Syracuse, contains, Onondaga

本文由国家自然科学基金 (61473294, 61370130, 61876198), 中央高校基本科研业务费专项资金(2015JBM033), 科学技术部国际科技合作计划 (K11F100010) 资助

收稿日期: 修回日期: 网络出版时间:

网络出版地址:

Lake”，远程监督方法会将包含上述两个命名实体的所有句子标注为 contains（包含）关系。

尽管远程监督是一种快速有效的训练语料标注手段，却存在严重的错误标注问题^[2]。例如，在句子“The Onondaga Nation is an 11-square-mile parcel in a valley south of Syracuse and about eight miles from the southern end of Onondaga Lake”中，实体“Syracuse”与实体“Onondaga Lake”之间并不具有“contains”关系，但仍会被看作是正例。对此，一些研究工作采用了多示例学习（Multi-instance Learning，简称 MIL）方法^[3]，并使用概率图模型来选择句子，进而解决错误标注问题^[4-5]。Zeng 等^[6]基于 at-one-leaf 假设，结合 MIL 和分段卷积神经网络（Piece-wise Convolutional Neural Networks，简称 PCNNs）来选择最有可能是正例的句子。Lin 等^[7]在 PCNNs 的基础上引入了句子级别的注意力机制，通过分配给句子不同的注意力权重来提取所有句子的有效特征。Ji 等^[8]提出了名为 APCNNs 的模型架构，同样使用了注意力机制，并加入了实体描述信息，在此类方法中取得目前最好效果。

虽然上述研究工作在远程监督关系抽取任务上取得了很好的效果，但仍存在两个问题。具体是：（1）句子中并非所有的单词都对关系标签的判断有贡献。例如，句子“The cultural appreciation of photography in France was driven by literary theorists like Roland Barthes and Jean Baudrillard , not by artists”描述了实体“Roland Barthes”与“France”之间的“nationality”关系，但子句“not by artists”对关系“nationality”的判定作用较小，从中提取的特征反而会会影响关系抽取模型的准确性；（2）训练过程中实体对的标签是不可变的（硬标签），这在一定程度上扩大了远程监督错误标注问题带来的影响。

针对以上问题，我们提出了一种新的融合门控机制的 PCNNs 模型（Gated PCNNs，简称 GPCNNs），利用门控机制^[9]对 PCNNs 提取的特征进行筛选，使其保留更多的与关系标签相关的特征。另外，我们在门控机制中，借鉴了软标签的思想^[10]，即标签信息会随着模型训练而改变。具体做法是，使用实体对之间的信息作为软标签，来指导对错误标签相关的特征的过滤，减弱训练过程中错误标签带来的影响。最后，结合句子级别的噪声过滤，获得包含较多正相关特征的句子。我们在公开数据集上进行测试，实验结果表明：我们提出的模型性能优于所有的基线系统，验证了所提出方法的有效性。

本文的组织结构如下：第 1 节介绍我们提出的 GPCNNs 模型架构和原理；第 2 节通过实验验证所提方法的有效性，并与已有方法进行比较；第 3 节介绍远程监督关系抽取任务的相关研究工作；最后对全文进行总结，并展望未来研究方向。

1 模型架构

给定一个包(Bag，包含相同实体对的句子集合) $B = \{s_1, s_2, \dots, s_n\}$ 和相应的两个实体，我们的模型将在 B 上预测各个实体关系标签的概率。我们提出的模型的整体架构如图 1 所示。

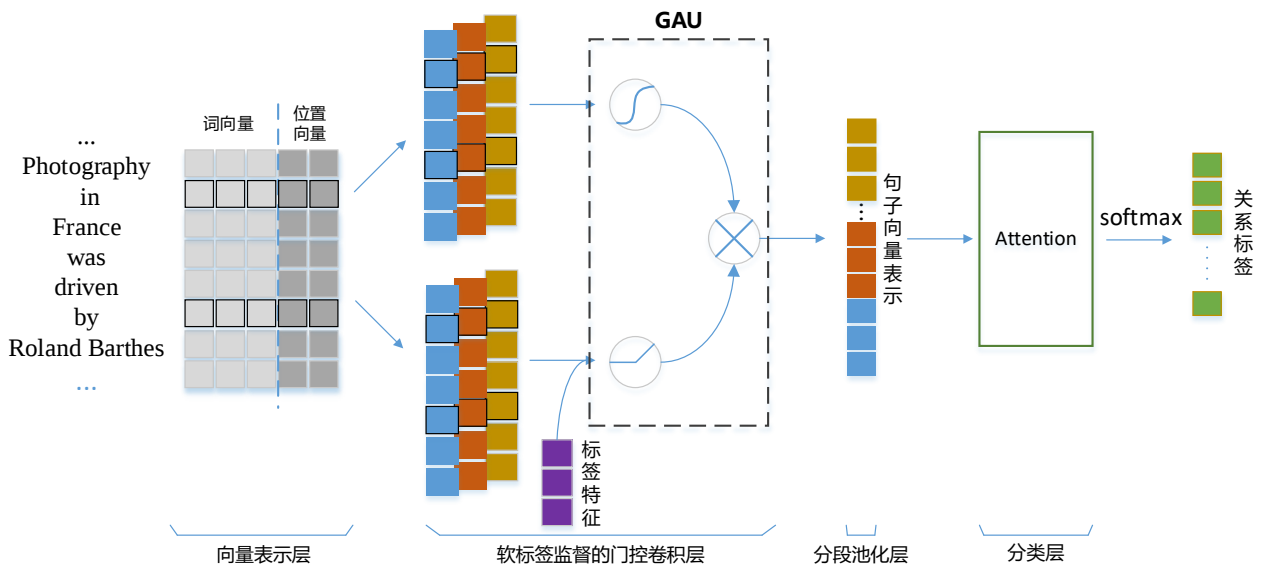


图 1 GPCNNs 整体架构图

Fig.1 The architecture of GPCNNs

图 1 中，给定一个句子 s ，首先我们将句子中的单词转换为低维向量，进而得到句子的矩阵表示（1.1 节向量表示层），之后使用门控 PCNNs 模块提取句子特征（1.2 节软标签监督的门控卷积层和 1.3 节分段池化层），最后通过注意力机制得到最终的各个标签得分（1.4 节决策层）。

相较于传统的 PCNNs 模型，我们对卷积层进行了改进，加入了门控机制（如图 1 软标签监督的门控卷积层所示），并在此基础上，引入了软标签的思想，使得模型可以更好地过滤词级别噪声。我们将在 1.2 节详细介绍门控机制和软标签。

1.1 向量表示层

词向量（Word Embedding）是单词的分布式表示，目的是将单词映射到一个低维的向量空间中。单词的向量表示通过查找预训练好的向量矩阵 $\mathbf{V}$ （或称查找表）获得，其中 $\mathbf{V} \in \mathbb{R}^{|\mathbf{V}| \times d^w}$ ， $|\mathbf{V}|$ 为矩阵 $\mathbf{V}$ 大小， d^w 是词向量的维度。

同已有研究工作一样，我们采用位置特征（Position Feature）来确定实体对位置，以便神经网络对单词与头实体（Head Entity） e_1 和尾实体（Tail Entity） e_2 之间的距离进行追踪。位置特征为当前单词到 e_1 和 e_2 的相对距离的组合，每个单词都有两个相对距离。图 2 为相对距离的一个例子，词“literary”到 e_1 （France）和 e_2 （Roland Barthes）的相对距离分别为 4 和 -3。随机初始化两个位置向量矩阵（ $\mathbf{PF}_1$ 和 $\mathbf{PF}_2$ ），我们便可通过它们将相对距离映射为实值向量，即位置向量（Position Embedding）。

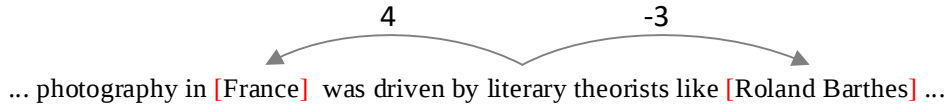


图 2 相对距离的例子

Fig.2 An example of relative distances

假设词向量的维度大小为 d^w ，且位置向量的维度大小为 d^p ，通过词向量和位置向量，我们将句子 s 转换为一个矩阵 $\mathbf{S} \in \mathbb{R}^{|\mathbf{s}| \times d}$ ，其中 $|\mathbf{s}|$ 为句子 s 的长度， $d = d^w + d^p$ 。随后将 $\mathbf{S}$ 作为神经网络的输入（如图 1 向量表示层所示）。

1.2 软标签监督的门控卷积层

卷积神经网络（Convolutional Neural Network，简称 CNN）可以有效提取输入矩阵的所有局部特征，并在此基础上进行全局预测。且 CNN 支持并行，计算效率高。因此，我们采用 CNN 来提取句子特征。

给定一个句子序列 $s = \{q_1, q_2, \dots, q_n\}$ ，其中 q_i 为第 i 个单词的向量表示， $q_i \in \mathbb{R}^d$ 。我们将卷积核大小设置为 w （图 1 所示的卷积层 $w = 3$ ），用 $\mathbf{S}_{ij}$ 表示 q_i 到 q_j 的拼接矩阵，相应的卷积操作权重矩阵为 $\mathbf{w} \in \mathbb{R}^{w \times d}$ 。则对 s 进行卷积操作后的结果为 $\mathbf{c} = \{c_1, c_2, \dots, c_{|\mathbf{s}| - w + 1}\}$ ：

$$c_j = f(\mathbf{w} \otimes \mathbf{S}_{(j-w+1):j} + b_c), \quad (1)$$

其中 $1 \leq j \leq |\mathbf{s}|$ ， $\otimes$ 为卷积运算， f 为非线性激活函数（如 $\tanh$ 函数）， $b_c \in \mathbb{R}$ 是偏置项。

1.2.1 软标签监督的门控机制

考虑到词级别噪声对模型性能的影响，我们采用门控机制来过滤词级别的噪声。基于 Gated Tanh Units（GTU）和 Gated Linear Units（GLU）的门控机制在语言模型中被广泛应用^[11-12]，且取得了良好效果。综合计算性能和有效性的考虑，我们对基于 GTU 的门控机制进行了改进，将其命名为 GAU（Gated Activation Units）模块（如图 1 所示），对应的运算结构为：

$$c_{\text{GAU}} = \tanh(\mathbf{w} \otimes \mathbf{S} + b_1) \times \text{relu}(\mathbf{v} \otimes \mathbf{S} + b_2), \quad (2)$$

其中 relu 函数会根据自身的输出来限制 $\tanh$ 函数的输出，以达到过滤句内噪声的目的。

在基于硬标签的方法中，实体对的关系标签（无论它们是否正确）在训练期间是不可变的，这在一定程度上扩大了远程监督错误标注问题对模型性能产生的负面影响，针对于此，我们在门控机制内引入了软标签策略，即训练时将标签替换为实体对之间的动态信息来指导特征过滤。

如图 1 所示, GAU 模块连接着两个卷积网络(一个为原始 CNN, 另一个带有标签特征)。我们使用两个实体间的双线性变换结果 $\mathbf{l}_{\text{relation}} = \mathbf{e}_1 \mathbf{W}_B \mathbf{e}_2$ ($\mathbf{W}_B$ 为模型参数) 作为实体对 $(\mathbf{e}_1, \mathbf{e}_2)$ 之间的关系标签, 指导模型在词级别上对不相关特征进行过滤。具体来说, 我们通过以下方式获得特征 $c_{\text{GAU}, j}$:

$$m_j = \text{relu}(\mathbf{w}_{m,j} \otimes \mathbf{S}_{(j-w+1):j} + \mathbf{l}_{\text{relation}} + b_m), \quad (3)$$

$$n_j = \tanh(\mathbf{w}_{n,j} \otimes \mathbf{S}_{(j-w+1):j} + b_n), \quad (4)$$

$$c_{\text{GAU}, j} = m_j \times n_j. \quad (5)$$

为了提取不同的特征, 通常需要使用不同的卷积核。因此, 我们采用了 n 个卷积核, 每次卷积操作都需要两个系数矩阵, 即 $\mathbf{W} = \{(\mathbf{w}_{m,1}, \mathbf{w}_{n,1}), (\mathbf{w}_{m,2}, \mathbf{w}_{n,2}), \dots, (\mathbf{w}_{m,n}, \mathbf{w}_{n,n})\}$, 则卷积后得到的最终特征表示 c_{ij} 为:

$$c_{ij} = c_{\text{GAU}, ij} = m_{i,j} \times n_{i,j} \quad (6)$$

其中 $1 \leq i \leq n$ 且 $1 \leq j \leq |s| - w + 1$ 。经过以上运算, 卷积层的整体输出结果为: $\mathbf{C} = \{c_1, c_2, \dots, c_n\}$ 。

1.3 分段池化层

最大池化操作通常被用于提取特征图 (Feature Map) 中最重要的特征, 但忽略了结构信息和细粒度信息。Zeng 等^[6]提出了一种新的池化机制, 根据给定的实体对, 将句子分为三部分后分别进行最大池化。对于卷积层结果 $\mathbf{C} \in \mathbb{R}^{|s|-w+1}$, 可以被划分为 $\mathbf{C} = \{c_{i,1}, c_{i,2}, c_{i,3}\}$, 则分段最大池化被定义为 $p_{ij} = \max(c_{i,j})$, 其中 $1 \leq i \leq n$, $j = 1, 2, 3$, 之后将所有的向量 $\mathbf{p}_i = [p_{i,1}, p_{i,2}, p_{i,3}]$ 进行拼接, 得到向量 $\mathbf{P} \in \mathbb{R}^{3n}$ (图 1 的分段池化层为 $n = 3$ 时的情况)。最后通过计算得到句子 s 的特征向量 $\mathbf{b}_s = \tanh(\mathbf{P})$ 。

1.4 决策层

已有的研究中, Lin 等^[7]和 Ji 等^[8]分别使用了双线性和非线性形式的注意力机制来弱化包中噪声句子对模型性能的影响。考虑到计算效率, 我们采用非线性形式的注意力机制。

我们同样使用 $\mathbf{l}_{\text{relation}} = \mathbf{e}_1 \mathbf{W}_B \mathbf{e}_2$ 作为标签信息。对于第 i 个包 B_i 中的第 j 个句子的特征向量 $\mathbf{b}_i^j$, 其注意力权重 α^j 的计算如下:

$$\alpha^j = \frac{\exp(\omega^j)}{\sum_k \exp(\omega^k)}, \quad (7)$$

$$\omega^j = \mathbf{W}_a^T (\tanh[\mathbf{b}_i^j; \mathbf{l}_{\text{relation}}]) + b_a, \quad (8)$$

则 B_i 的最终向量表示为 $\mathbf{r}_i = \sum_{j=1}^{|B_i|} \alpha^j \mathbf{b}_i^j$ 。

然后包的向量表示被送入 softmax 分类器以预测最终的关系标签并计算交叉熵目标函数:

$$J(\theta) = -\sum_{i=1}^T \log p(y_i | \mathbf{r}_i; \theta). \quad (9)$$

2 实验

2.1 数据集及评测指标

我们采用 Riedel 等^[2]公布的数据集来评测我们的模型, 该数据集由 Freebase 中的关系事实 (Relation Facts) 和 New York Times (NYT) 语料库对齐生成。其中, 我们将对齐后的 2005-2006 期间的句子作为训练集, 2007 年的句子作为测试集。该数据集共有包含 NA 标签 (指实体对之间不存在关系) 在内的 53 种关系标签。训练数据共包含 522611 个句子、281270 个实体对和 18252 个关系事实; 测试数据共包含 172448 个句子、96678 个实体对以及 1950 个关系事实。

我们使用 held-out evaluation 方法来评估我们的模型。该方式提供了一种近似的准确率 (Precision) 测量

方法，即分别从测试集和 Freebase 中发现关系事实，并进行比较来评估模型，无需费时的人工评估。我们将在实验中展示总体 P-R 曲线（Precision/recall Curve）和 Precision@N（Precision at Top N Predictions）。

2.2 实验设置

已有研究结果表明使用预训练好的词向量可以提高模型收敛速度，提升模型性能。因此，本文使用了 word2vec 工具^[13]在 NYT 语料库上训练词向量，并使用出现频率 100 以上的单词来构建词汇表。

实验过程中，我们在训练集上使用 3 折交叉验证来选择模型参数。卷积核大小（Window Size）取值范围为{3, 5, 7}，卷积核数量（Feature Maps）范围为{50, 100, 200, 230}，batch size 取值范围为{50, 100, 160}，同时采用了 dropout 策略和 Adadelta 优化算法来训练模型，相应学习率（Learning Rate）取值范围为{0.001, 0.01, 0.1, 0.5}。最终获得的各项超参数如下：卷积核大小为 3，卷积核数量为 230，batch size 为 50，词向量维度和位置向量维度分别为 50 和 5，学习率为 0.1，dropout 比例为 0.5。

2.3 模型整体性能评测

我们将提出的 GPCNNs 模型同三个经典模型进行对比：PCNNs+MIL^[6]，PCNNs+ATT^[7]，以及 APCNNs^[8]。其中 PCNNs+MIL 会选择包中得分最高的句子作为包的表示；PCNNs+ATT 和 APCNNs 分别使用了句子级别的双线性和非线性形式的注意力机制，综合包内所有句子信息作为包的表示。另外，我们将 GAU 添加到了 PCNNs+MIL 模型中（名为 PCNNs+MIL+GAU）来验证 GAU 的有效性；同时我们把 GAU 替换成 GTU（名为 PCNNs+GTU）来验证实体对的双线性变换作为软标签的有效性。图 2 展示了我们提出的模型和所有基线模型的总体 P-R 曲线，表 1 显示了 N 分别为 100, 200 和 300 时的 Precision@N 值：

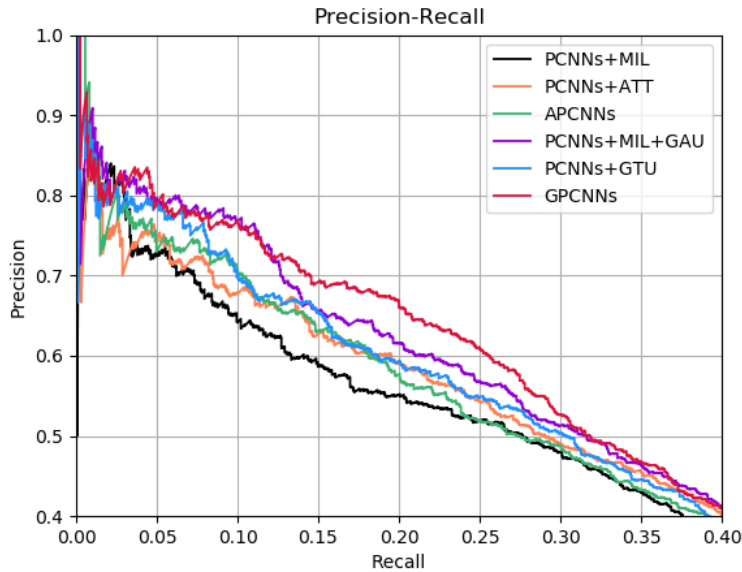


图 2 各个模型的总体 P-R 曲线

Fig.2 Aggregate precision/recall curves for a variety of models

表 1 PCNN+MIL, PCNN+ATT, APCNNs, GPCNNs 的 Precision@N

Table 1 Precision@N of PCNN+MIL, PCNN+ATT, APCNNs, GPCNNs

Precision@N(%)	Top 100	Top 200	Top 300	Average
PCNNs+MIL	72.89	69.23	64.05	68.72
PCNNs+ATT	74.26	72.14	68.44	72.61
APCNNs	76.24	74.13	69.44	73.27
PCNNs+MIL+GAU	81.19	77.61	74.42	77.74
PCNNs+GTU	78.22	76.62	68.77	74.54
GPCNNs	83.16	77.60	74.44	78.13

通过图 2 和表 1 可以看出：

- 1) 所有加入 GAU 的模型预测结果都优于不加 GAU 的基线模型，表现为相同召回率的情况下，加入 GAU 的模型取得了更高的准确率。这说明门控机制结合注意力机制相较于单纯的注意力机制可以过滤掉更多的负相关特征，验证了 GAU 模块的有效性。
- 2) GPCNNs 的整体 P-R 曲线相较于其他模型下降更为平缓，曲线下覆盖面积最大；Precision@N 值皆达到了最高，平均下来较基线模型提高了 4% 左右。这说明了 GPCNNs 性能明显优于所有的基线模型，说明了我们模型的整体有效性。
- 3) GPCNNs 相较于 PCNNs+GTU 取得了更好的效果。这说明了双线性变换可以有效映射实体对间的关系特征，结合门控机制可以有效地提取词级别的特征；同时也验证了软标签策略可以更为有效地筛选特征。

2.4 门控机制性能评测

为了更直观地验证门控机制的有效性，我们通过以下两种方式去除注意力机制并进行验证：1) 去除 GPCNNs 的注意力选择机制（命名为 PCNNs+GAU），即在句子级别上进行关系抽取任务；2) 将 GAU 单元应用于 CNNs（我们将其命名为 CNNs+GAU），并将 CNNs 和 PCNNs 作为基线模型设置对比实验。实验结果如下图 3 所示：

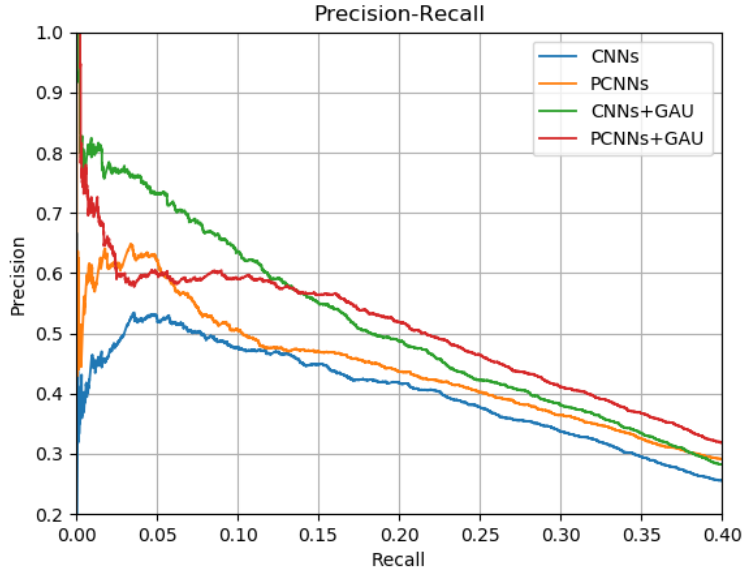


图 3 句子级别抽取中各个模型的总体 P-R 曲线

Fig.3 Aggregate precision/recall curves for a variety of models on the sentence-level extraction

从图 3 可以看出：

- 1) PCNNs+GAU 和 CNNs+GAU 相较于 PCNNs 和 CNNs 都取得了显著的提升，说明在不同的池化机制下 GAU 模块都有效提高模型整体性能，验证了门控机制的有效性，同时也反映了 GAU 模块的鲁棒性。
- 2) PCNNs+GAU 的整体曲线最为平缓，且 P-R 曲线下覆盖面积最大，说明了词级别噪声会对关系抽取模型的性能产生较大影响，验证了 GAU 模块的有效性，同时表明 GAU 模块可以有效地过滤词级别噪声，提高远程监督关系抽取模型的整体性能。

2.5 样例分析

为了显式证明词级别词级别噪声对模型性能的影响以及我们提出的模型在词级别噪声过滤上的有效性，我们展示了测试过程中包内句子注意力权重分配的一个例子。如表 2 所示，该包共有 3 个句子，这些句子均含有短语搭配 “president emeritus of”，该短语可以有效表明实体 “John Brademas” 和 “New York

University”之间的“/business/person/company”关系，故3个句子均为正例。

可以看出，表2的句子中除了短语搭配“president emeritus of”之外，其他单词与实体对关系“/business/person/company”并不具有直接或间接的联系，属于噪声。且相对于其他句子来说，第1个句子包含的噪声最多，第3个句子次之，第2句最少。由于缺乏有效的词级别噪声过滤机制，APCNNs模型分配给了第2句较高的权重（0.665），第1句和第3句较少的权重（0.085,0.249），而我们提出的GPCNNs模型则基本不受噪声的影响，分配给了三个句子相近的权重（0.379,0.313,0.308）。注意力权重的分配情况验证了我们提出的模型可以有效过滤词级别噪声1717，更充分地利用包中的监督信息。

表2 注意力权重 α 的例子

Table 2 An example of Weight α

实体关系对	句子	APCNNs	GPCNNs
(John Brademas, New York University, /business/person/company)	1. Thirty-five years ago, President vetoed the legislation, refusing to encourage "the family-centered child rearing". "I don't think we've ever recovered from that veto message." said John Brademas , <u>president emeritus of New York University</u> , and, as a former Democratic congressman from Illinois, a sponsor of that legislation.	0.085	0.379
	2. An article on Jan.11 about a conference in New York misidentified the home state of John Brademas , <u>president emeritus of New York University</u> .	0.665	0.313
	3. Correction: January 25, 2006, Wednesday An article on Jan.11 about a conference in New York misidentified the home state of John Brademas , <u>president emeritus of New York University</u> .	0.249	0.308

3 相关工作

由于人工标注成本高，远距离监督方法在关系抽取任务中发挥着越来越重要的作用。然而，这种方式面临着标签错误标注所带来的挑战，容易产生大量的噪声实例。对此，Riedel等^[2]引入了多示例学习，将远程监督关系抽取建模为多示例单标签问题，在包级别上预测标签。后续的一些研究工作^[4-5]则采用了多示例多标签学习，并使用概率图模型来选择句子。但是，以上方法都严重依赖于NLP工具生成的特征质量，深受错误传递问题的困扰。Zeng等^[6]提出了使用PCNNs来捕捉句子结构信息和特征，结合MIL来选择最有可能是正例的句子。但由于只使用包内的一个句子，使得MIL不能充分利用监督信息。在此基础上，一些研究工作^[7-8]提出了分别使用双线性和非线性的注意力机制来选择句子（PCNNs+ATT, APCNNs），使用包内句子向量的权重加和来作为包的向量表示，使用了包内所有句子的监督信息。然而，上述工作均是在句子级别上过滤噪声数据，忽略了词级别的噪声过滤，导致不能充分地利用包中的监督信息。另一方面，训练过程中实体对的固定关系标签（硬标签）也在一定程度上扩大了错误标注问题带来的影响。

不同于已有的研究工作，我们提出了GPCNNs模型，利用门控机制对PCNNs提取的特征进行噪声过滤，即在词级别上过滤噪声。此外，我们还引入了软标签的思想^[10-12]，在门控机制中用实体对的双线性变换结果作为句子的实体标签来过滤噪声，使得过滤结果更为精确。实验结果验证了GPCNNs模型的有效性。另外，Liu等^[14]也提出了一种词级别噪声过滤的方法。我们提出的模型和他们的不同之处在于：1）Liu等使用了NLP工具建立依存子树，并使用迁移学习引入外部先验知识来进行词级别噪声过滤，我们提出的模型仅使用了门控机制，并没有使用NLP工具和外部知识；2）我们提出的模型引入了软标签思想，减弱了错误标签对词级别噪声过滤的影响，进一步提高了词级别噪声过滤的精度。

此外，最近强化学习方法也被用于远程监督关系抽取任务上^[15-16]，在模型训练之前使用相应方法对语料库中的句子进行筛选，并在此方面取得了较好的效果。而我们的模型主要集中在模型训练过程中词级别的噪声过滤方面。

4 结语

针对词级别的噪声过滤问题，本文提出了一种新的融合门控过滤机制的分段池化卷积神经网络方法，将其应用于远程监督关系抽取任务。门控机制可以有效过滤词级别噪声，结合句子级别的筛选机制，显著提升了远程监督关系抽取模型的整体性能。另外，我们在门控机制中引入了软标签思想，使用实体对之间的双线性变换来指导特征过滤，提高词级别噪声过滤的精度。实验结果表明，我们的模型优于所有的基线模型，达到了最好的效果。今后，我们将结合强化学习，引入先验知识，探索进一步提高抽取性能的方法。

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 Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009: 1003-1011.
- [2] Riedel S, Yao L, McCallum A. Modeling relations and their mentions without labeled text[C]. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2010: 148-163.
- [3] 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.
- [4] 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-Volume 1. Association for Computational Linguistics, 2011: 541-550.
- [5] Surdeanu M, Tibshirani J, Nallapati R, et al. Multi-instance multi-label learning for relation extraction[C]. Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning. Association for Computational Linguistics, 2012: 455-465.
- [6] Zeng Daojian, Liu Kang, Chen Yubo, 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, 2015: 1753-1762.
- [7] Lin Yankai, Shen Shiqi, Liu Zhiyuan, et al. Neural relation extraction with selective attention over instances[C]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016, 1: 2124-2133.
- [8] Ji Guoliang, Liu Kang, He Shizhu, et al. Distant supervision for relation extraction with sentence-level attention and entity descriptions[C]. Thirty-First AAAI Conference on Artificial Intelligence, 2017.
- [9] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [10] Wang Guanying, Zhang Wen, Wang Ruoxu, et al. Label-free distant supervision for relation extraction via knowledge graph embedding[C]. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018: 2246-2255.
- [11] Kalchbrenner N, Espeholt L, Simonyan K, et al. Neural machine translation in linear time[J]. arXiv preprint arXiv:1610.10099, 2016.
- [12] Dauphin Y N, Fan A, Auli M, et al. Language modeling with gated convolutional networks[C]. Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017: 933-941.
- [13] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [14] Liu Tianyu, Zhang Xinsong, Zhou Wanhao, et al. Neural relation extraction via inner-sentence noise reduction and transfer learning[J]. arXiv preprint arXiv:1808.06738, 2018.
- [15] Feng Jun, Huang Minlie, Zhao Li, et al. Reinforcement learning for relation classification from noisy data[C]. Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
- [16] Qin Pengda, Xu Weiran, Wang W Y. Robust distant supervision relation extraction via deep reinforcement learning[J]. arXiv preprint arXiv:1805.09927, 2018.