

基于集成学习方法的实体关系抽取^①



丰小丽, 张英俊, 谢斌红, 赵红燕

(太原科技大学 计算机科学与技术学院, 太原 030024)

通讯作者: 丰小丽, E-mail: 1334852927@qq.com

摘要: 基于神经网络的实体关系抽取模型已经被证明了它的有效性, 但使用单一的神经网络模型在不同的输入条件下, 会表现出不同的结果, 性能不太稳定. 因此本文提出一种利用集成学习思想将多个单一模型集成为一个综合模型的方法. 该方法主要使用 MLP (MultiLayer Perceptron) 将两个单一模型 Bi-LSTM (Bi-directional Long Short-Term Memory) 和 CNN (Convolutional Neural Network) 集成为一个综合模型, 该模型不仅可以充分利用两个单一模型的优势, 而且可以利用 MLP 的自学习能力与自动分配权重的优势. 本研究在 SemEval 2010 Task 8 数据集上取得了 87.7% 的 $F1$ 值, 该结果优于其他主流的实体关系抽取模型.

关键词: 实体关系抽取; Bi-LSTM; CNN; 集成学习; MLP

引用格式: 丰小丽, 张英俊, 谢斌红, 赵红燕. 基于集成学习方法的实体关系抽取. 计算机系统应用, 2021, 30(6):255-261. <http://www.c-s-a.org.cn/1003-3254/7952.html>

Entity Relation Extraction Based on Ensemble Learning Method

FENG Xiao-Li, ZHANG Ying-Jun, XIE Bin-Hong, ZHAO Hong-Yan

(School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China)

Abstract: The entity relation extraction model based on neural networks has been proven effective, but a single neural network model is unstable because it can yield various results with different inputs. Therefore, this study proposes a method to integrate multiple single models into a comprehensive one using the idea of ensemble learning. Specifically, this method integrates Bi-directional Long Short-Term Memory (Bi-LSTM) and Convolutional Neural Network (CNN) into a comprehensive model through MultiLayer Perceptron (MLP), which cannot only fully take advantage of the two single models, but also make use of the self-learning ability and automatic weight allocation of MLP. This study obtains $F1$ of 87.7% on the SemEval 2010 Task 8 dataset, which is better than other mainstream entity relation extraction models.

Key words: entity relation extraction; Bi-LSTM; CNN; ensemble learning; MLP

信息抽取旨在从大规模半结构化或非结构化的自然文本中抽取结构化数据. 实体关系抽取作为该领域重要的一部分, 其目的是抽取各实体之间的语义关系. 准确的关系分类有利于精确的句子解释和文本的语义理解, 从而更有效地进行自然语言处理 (Natural Language Processing, NLP) 任务. 此外, 关系抽取在问答系统、机器翻译和信息检索等方面有着重要的应用价值和现实意义.

目前实体关系抽取任务均是在给定句子实体的情况下, 在单句内和限定关系的集合中进行实体关系抽取. 在该场景下, 深度学习方法应用较为广泛, 而且效果较好. 与传统方法相比, 深度学习方法能够捕获句子更深层次的语义信息. Socher 等^[1] 在实体关系抽取任务中使用循环神经网络 (Recurrent Neural Networks, RNN), 将句子的句法结构考虑在内, 但却忽略了单词之间的位置信息; Zhang 等^[2] 使用 RNN 进行实体关系抽

① 收稿时间: 2020-09-26; 修改时间: 2020-11-05; 采用时间: 2020-11-17; csa 在线出版时间: 2021-06-01

取时虽然引入了位置信息,但梯度消失问题使其很难解决远距离依赖问题;Zeng等^[3]在基于位置信息的基础上使用卷积神经网络(Convolutional Neural Networks, CNN),虽然CNN可以利用多个n-gram特征,使提取的特征更丰富,但由于卷积核通常不会太大,因此也不能很好地学习到句子中的远程语义信息.Miwa和Bansal^[4]在句子序列上使用基于句法树的长短期记忆网络(Long Short-Term Memory, LSTM)来获取句子序列的上下文信息,克服了RNN梯度消失的问题,解决了长距离依赖的问题.然而,这些模型都只是在某一方面表现更好,而且其性能不是很稳定^[5],不能有效的解决关系抽取中的噪声数据和错误标注的数据.

针对上述问题,本文提出一种新的集成(ensemble)神经网络模型.该模型主要使用两个不同的单一模型(也称为弱分类器)——双向长短期记忆网络(Bi-directional LSTM, Bi-LSTM)和CNN,然后将其集成到使用多层感知机(MultiLayer Perceptron, MLP)的综合模型中.在综合模型中利用Bi-LSTM解决句子长距离依赖问题的优势,利用CNN解决并行计算问题和充分提取特征信息的优势,利用MLP高度并行性和良好容错性与自适应性的优势,自动的学习出实体对的语义关系.Rokach等^[6]证明使用某些策略将多个模型结合起来,可以有效地提升机器学习的性能.

常用集成学习方法的权重分配策略主要有简单平均法和加权平均法,虽然这两种方法较为简单,但其不能拟合非线性数据,而神经网络具有良好的自学习和充分逼近复杂的非线性关系的能力.因此,本文使用人工神经网络进行权重分配;此外,决策树与随机森林等方法,虽然有一定的效果,但其缺点是容易产生过拟合,而对于神经网络则可通过参数正则化、Dropout以及选择合适的网络结构等方式来防止该现象的发生.

1 相关工作

实体关系抽取作为NLP一项基础且重要的任务,受到了广泛的关注和研究.早期,基于规则^[7]的关系抽取方法主要依赖领域专家构建大规模的模式匹配规则,费时费力,且领域自适应性较差,因此发展受到限制;基于特征^[8]的方法,通过提取文本中的重要特征表征实体之间的关系,如依存句法树^[9]等,但该方法依赖于NLP工具的准确性,因此,可能会导致错误传播等问题;基于核函数的方法,主要通过核函数来计算不同关

系实例之间的相似度,且核函数的设计需要大量人员进行研究,因此其不适用于大规模语料的关系抽取任务.

目前,基于神经网络的方法应用较为广泛.常用的主要包括基于CNN和RNN两种网络模型.

基于卷积神经网络的方法主要有:Chen等^[10]提出基于强化学习的句级标签去噪模型,并利用PCNN(Piecewise Convolutional Neural Network)作为句子编码器来进行远程监督关系抽取;曹春萍等^[11]使用ATT-CNN与BSRU(简单循环神经网络)结合,解决了生物医学中化学物质与疾病的关系;Lin等^[12]利用CNN结合注意力机制进行远程监督关系抽取,缓解标注语料较少的问题;Liu等^[13]利用正确标注的实体对的语义信息,提出一种实体对级的去噪方法,该方法可以在训练过程中动态的纠正错误标签;Qin等^[14]则是将生成对抗网络和CNN进行结合来缓解远程监督关系抽取中的噪声问题;刘伟等^[15]使用语法树与CNN构建的Tree-base CNN,解决CNN对语法信息编码不足的问题.基于循环神经网络的方法主要有:Gupta等^[16]通过循环神经网络对最短和增强的依赖路径进行建模,以提取句子内部和跨句子边界的关系;李卫疆等^[17]使用Bi-LSTM解决单词之间的依赖关系,而且将位置信息、词汇、句法和语义信息相融合,再使用自注意力机制重点学习序列内部的特征信息;Phi等^[18]使用两层双向门控循环单元(Bi-directional Gated Recurrent Unit, BGRU)提取句子上下文信息,同时添加了分段注意力机制,以突出每个片段对句子的重要性.

卷积神经网络与循环神经网络在实体关系抽取任务中表现各异:卷积神经网络的优势在于其关注关系实例的局部信息,缺点是忽视了序列的全局信息;循环神经网络则相反,可较好地捕获序列的整体特征,但对局部特征的抽取较为欠缺.因此,就学习层面上二者互为补充.所以本文使用集成学习的思想,充分利用二者的优势,将两者集成为一个综合模型,该模型不仅具有卷积神经网络关注局部信息的特征,而且具有循环神经网络关注全局信息的优势,同时还可以利用MLP进行权重分配.

另外,集成学习是一种著名的机器学习范例,Peng等^[19]利用集成学习方法进行化学蛋白质的关系抽取;Yang等^[5]在进行远程监督关系抽取时,提出了一种集合神经网络模型的自适应增强LSTM模型;Li等^[20]也表明集成学习技术可以成功的应用于迁移学习任务.

同样, 本文所提方法也是一种通用的方法, 该方法在其他类似任务中同样适用。

2 基于集成学习方法的实体关系抽取模型

本文提出的综合模型结构如图1所示, 该模型主要由以下5个部分组成。

(1) 输入层: 将原始语料中有格式错误、内容或逻辑错误等的数据进行清洗, 保留清洗后的特征数据;

(2) 词嵌入层: 将(1)中清洗后句子中的每个单词映射成低维稠密向量, 并与单词位置向量拼接后作为弱分类器的输入;

(3) 弱分类器层: 将(2)中词向量与位置向量拼接后的向量输入到弱分类器 Bi-LSTM 和 CNN 中, 分别进行特征提取与分类, 保存其输出结果;

(4) 强分类器层: 将(3)中弱分类器输出结果 Output_BiLSTM 和 Output_CNN 拼接起来, 输入到强分类器 MLP 中, 以得到句子的高维语义信息;

(5) 输出层: 利用 Softmax 函数进行实体关系进行分类。

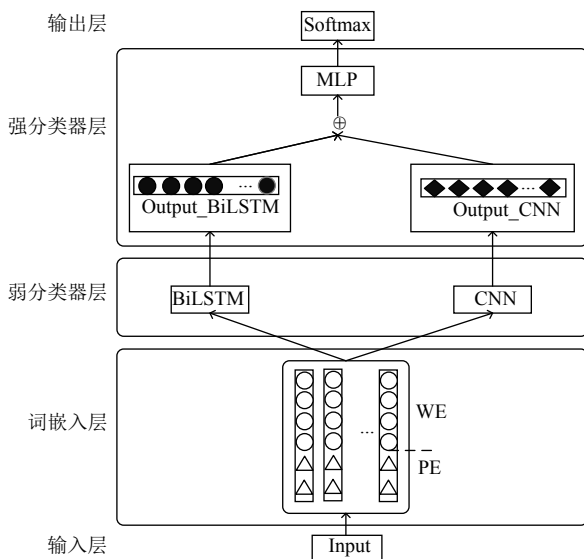


图1 综合模型结构图

2.1 词向量表示

使用词向量表示的目的是将输入的句子转换成向量的形式, 方便计算机对其进行读取。为了充分利用单词在句子中的语义信息, 本文将每个输入的单词转换成词嵌入和位置嵌入的结合。

(1) 词向量

词嵌入 (Word Embeddings, WE) 目的是把单词转

换成低维密集向量, 以捕获单词的句法和语法特性。假设一个由 n 个词组成的句子 $S = \{w_1, w_2, \dots, w_n\}$, 可以通过词嵌入矩阵 $M \in R^{d \times V}$, 将每个词 w_i 映射为实值向量 e_i , 其中, V 表示词表大小, d 表示词向量维度, 因此, 句子的词向量可表示为 $S_w = (e_1, e_2, \dots, e_n) \in R^{d \times n}$ 。

(2) 位置向量

位置嵌入 (Position Embedding, PE)^[3] 用于表征句子中某个单词与两个实体 e_1, e_2 的相对位置。如句子: The $\langle e_1 \rangle$ deficits $\langle e_1 \rangle$ are caused by $\langle e_2 \rangle$ people $\langle e_2 \rangle$ save too much of their money. 单词 caused 相对于 e_1 (deficits) 的距离为 2, 相对于 e_2 (people) 的距离为 -2。即若实体的位置为 0, 则实体左侧单词相对于该实体的 1, 2, ... 将计算出来的相对距离经过位置嵌入矩阵 $M' \in R^{c \times L}$ 转换为实值向量 p_i , 其中, c 是位置向量维度, L 是句子中任何一个单词相对于两个实体的最大距离; 由于每个句子有两个标记实体, 所以每个词对应两个位置向量, 因此, 第 i 个单词的位置向量表示如式 (1)。

$$p_i = p1_i \oplus p2_i \quad (1)$$

其中 $p1_i, p2_i$ 是第 i 个单词相对于 e_1, e_2 的位置向量表示。因此, 句子的位置向量表示为 $S_p = (p1, p2, \dots, pn) \in R^{2c \times n}$ 。

最后将词向量和位置向量拼接起来, 得到最终的句子向量表示 $U = S_w \oplus S_p \in R^{n(d+2c)}$ 。

2.2 Bi-LSTM 弱分类器模型

为了得到句子的高维语义信息, 并捕获序列的全局特征, 本文使用 Bi-LSTM 模型对输入的句子序列进行特征提取, 其结构如图2所示。由于 Bi-LSTM 可以从序列的前向和后向同时进行语义捕捉, 因此其可以充分利用输入序列的上下文信息来提取特征信息。图2中词嵌入层是单词的词向量和位置向量拼接后的特征向量 $U = (u_1, u_2, \dots, u_n)$, 其中, $u_i = e_i + p_i, u_i \in R^{d+2c}$ 。Bi-LSTM 模型隐藏层的计算方式如式 (3) 和式 (4)。

$$\vec{h} = lstm(u_t, \vec{h}_{t-1}) \quad (2)$$

$$\overleftarrow{h} = lstm(u_t, \overleftarrow{h}_{t+1}) \quad (3)$$

其中, u_t 为 t 时刻的输入向量, \vec{h} 为前向的隐藏状态, \overleftarrow{h} 为后向的隐藏状态, \vec{h}_{t-1} 表示前一时刻的隐藏状态, \overleftarrow{h}_{t+1} 表示下一时刻的隐藏状态。所以, t 时刻的隐藏状态为前向和后向隐藏状态的拼接, 其表示方式如式 (5)。

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t \quad (4)$$

因此最终隐藏层输出可表示 $H = (h_1, h_2, \dots, h_n)$.

由于实体关系抽取任务属于多分类问题 $y \in \{1, \dots, m\}$, m 表示关系类别个数, 因此经过 Softmax 分类器, 得到每个关系类别的预测概率 \hat{y}_1 :

$$\hat{y}_1 = \text{softmax}(H) \quad (5)$$

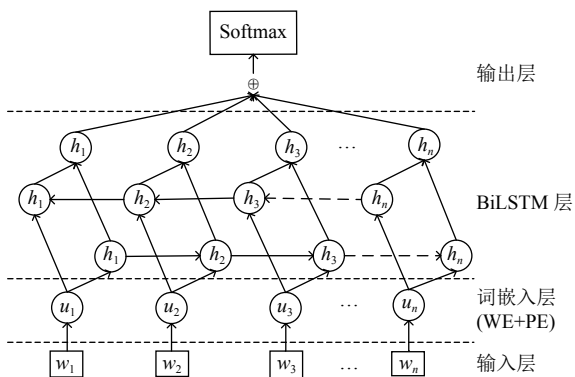


图2 Bi-LSTM弱分类器模型结构图

2.3 CNN弱分类器模型

CNN 可以解决实体关系抽取任务中并行计算的问题, 同时增加网络层数可以解决长距离依赖的问题, 但随着层数的增加, 模型的参数就会增多, 从而导致训练时间过长. 所以本文选用 2 层卷积层和 2 层全连接层进行关系抽取, 既保证并行性又防止训练时间过长, 还可以提取到更有效的特征. 此外, 为了防止由原始数据中句子长度不同而引起的输入矩阵维度不同的问题, 本文将数据集中的每条句子做 padding 操作, 将其 padding 到句子的最大长度 Max_len, 同时为了保证输出层和输入层维度的统一, 使用 SAME 做本次 padding 操作. 本文使用的 CNN 网络结构如图 3 所示.

在该模型中, 输入到卷积层的是 $U = (u_1, u_2, \dots, u_n)$, 若滑动窗口大小为 k , 当窗口在 U 上滑动时, 向量 $u_i, u_{i+1}, \dots, u_{i+k-1}$ 落到第 j 个窗口的向量表示如式 (7):

$$Q_j = (u_j, u_{j+1}, \dots, u_{j+k-1}) \quad (6)$$

若卷积核为 $W \in R^{n(d+2c)}$, c 为卷积核的宽度, 则向量 Q_j 经过卷积层与池化层输出的特征信息可使用式 (8) 和式 (9) 计算.

$$x_j = f(Q_j \odot W + b) \quad (7)$$

$$v_j = \max(x_j) \quad (8)$$

$$v = [v_1, v_2, \dots, v_s] \quad (9)$$

其中, \odot 是卷积操作, b 是偏置, f 是非线性激活函数, s 是卷积核的数目. 最后经过全连接层与 Softmax 分类器得到关系预测概率, 其计算方式如式 (11) 和式 (12).

$$z = f(v) \quad (10)$$

$$\hat{y}_2 = \text{softmax}(z) \quad (11)$$

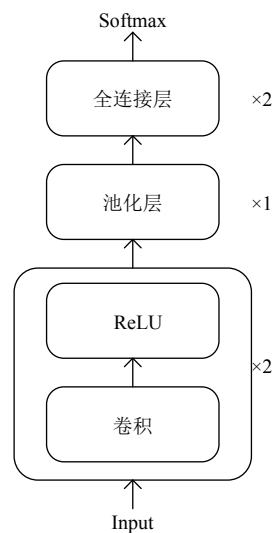


图3 CNN弱分类器网络结构图

2.4 综合模型

集成学习是将一系列相对较弱的模型以某种恰当的方式组合起来, 得到比使用单个模型更鲁棒的强模型. 本文使用 MLP 进行模型集成, 其网络模型结构如图 4 所示. 在该模型中, 输入主要包括两个部分: Output_BiLSTM、Output_CNN. 其中, Output_BiLSTM 是 Bi-LSTM 弱分类器计算的关系类别预测概率 \hat{y}_1 , 取其最大值 $\max(\hat{y}_1)$ 作为关系类别标签, 经过独热编码的形式将其转换为 m 维的关系矩阵; Output_CNN 的表示方法与 Output_BiLSTM 同理. 因此, 综合模型的输入可表示为式 (13):

$$X = \hat{y}_1 \oplus \hat{y}_2 \quad (12)$$

实验过程中若 MLP 选择较少的网络层数, 则其特征提取的效果较差, 随着网络层数的不断增加, 模型复杂度增加, 虽然其提取的效果有所提高, 但网络参数也会不断增多, 因此会导致模型训练效率低下的结果. 所以本次实验选择 3 层隐藏层的 MLP 进行特征提取, 在提高训练效率的同时, 保证实体关系抽取的性能. 使用 MLP 进行信息传播可以使用式 (14) 和式 (15) 进行计算:

$$z^{(l)} = W^{(l)} * a^{(l-1)} + b^{(l)} \quad (13)$$

$$a^{(l)} = f_i(z^{(l)}) \quad (14)$$

其中, $z^{(l)}$ 表示 l 层神经元的净输入, $a^{(l)}$ 表示 l 层神经元的输出, $f_i(\cdot)$ 表示神经元的激活函数, $W^{(l)}$ 表示 $l-1$ 层到 l 层的权重矩阵, $b^{(l)}$ 表示 $l-1$ 到 l 层的偏置. 最后利用 Softmax 分类器, 得到最终每个关系类别的预测概率 $\hat{y} = softmax(z^{(l)})$.

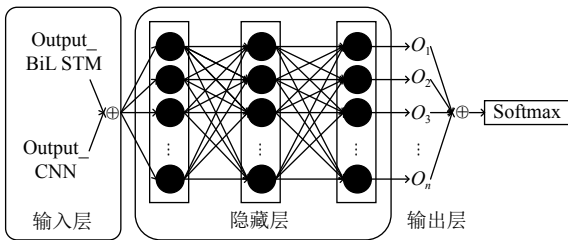


图4 使用 MLP 进行集成时的网络结构图

3 相关实验与结果分析

3.1 数据集

为了评估综合模型在实体关系抽取任务中的有效性, 本文使用实体关系抽取公开数据集 SemEval 2010 Task 8 进行验证. 在该数据集中, 总共包含 10717 个标记样本, 其中用于训练的样本数量有 8000 个, 用于测试的样本数量有 2717 个. 但为了产生多样性大的弱学习器, 本文首先选择两个不同的弱学习器进行集成学习, 其次使用数据样本对其进行扰动, 即随机选取原样本的 80% 作为训练数据, 剩下的 20% 作为测试数据, 因此, 训练样本数为 8573, 测试样本数为 2144. 该样本总共包含 9 种具体的关系类型和一种“other”类别的关系. 其具体关系类型如表 1 所示.

表1 关系类型

关系标签	关系类型	描述
0	Cause-Effect	因果关系
1	Component-Whole	部分与整体关系
2	Content-Container	内容与包含关系
3	Entity-Destination	实体与目的关系
4	Entity-Origin	实体与来源关系
5	Instrument-Agency	工具与机构关系
6	Member-Collection	成员与组织关系
7	Message-Topic	信息与主题关系
8	Product-Producer	产品与生产者关系
9	other	其他关系

3.2 评价指标

通常情况下, 实体关系抽取任务的评价指标主要

包括准确率 (*precision*)、召回率 (*recall*) 以及 *F1* 值. 对于每种关系类型, 主要有以下 4 种判断: *TP* (True Position)、*FP* (False Position)、*TN* (True Negative) 和 *FN* (False Negative), 它们分别表示将正类预测为正类、将负类预测为正类、将正类预测为负类和将负类预测为负类. 所以这 3 种评价指标的计算方式如式 (16)–式 (18).

$$precision = \frac{TP}{TP + FP} \quad (15)$$

$$recall = \frac{TP}{TP + FN} \quad (16)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (17)$$

上面的计算公式是针对单个关系类别进行的评估, 为了在本文使用的数据集中进行验证, 使用每种关系类型的宏平均 (*macro average*) 验证本文的实验结果, 即对每一种关系类别得到的概率求和后取平均. 若其关系类别的个数用 m 表示, 则具体的计算方式如式 (19)–式 (21).

$$P = \frac{1}{m} \sum_{i=1}^m precision(i) \quad (18)$$

$$R = \frac{1}{m} \sum_{i=1}^m recall(i) \quad (19)$$

$$F1 = \frac{1}{m} \sum_{i=1}^m F1(i) \quad (20)$$

3.3 参数设置

本实验中使用的参数设置主要根据经验设定, 如词嵌入使用预训练好的 Glove300 维向量等; 还有一部分根据实验的效果进行参数调整, 如模型训练的 *batch_size* 设定等, 其具体的实验参数设置如表 2 所示.

表2 实验参数设置

参数	值
Word dimension (WE)	300
Position dimension (PE)	100
Dropout	0.7
Max_len	100
LSTM unit hidden number	128
LSTM batch_size	128
CNN kernel_size	2,3
CNN filter number	100
MLP batch_size	128

3.4 实验结果及分析

为了验证本文提出的综合模型对实体关系抽取性

能的影响,主要进行以下对比实验:

(1) 验证综合模型与单一模型对实验效果的影响

综合模型与两个单一模型对实验结果的影响如表3所示.该结果表明,综合模型 Ensemble 相较于单一模型 Bi-LSTM 在 $F1$ 值上提高了 21.97%,相较于单一模型 CNN 在 $F1$ 值上提高了 28.92%.该结果充分说明综合模型可以利用弱分类器的优势,得到比使用单个弱分类器进行实体关系抽取时更好的结果.

表3 综合模型与单个模型的实验对比

模型	P	R	$F1$
Bi-LSTM	66.73	65.10	65.73
CNN	60.12	58.10	58.78
Ensemble	87.77	88.21	87.70

(2) 综合模型与主流模型的对比

为了验证综合模型在实体关系抽取任务中的有效性,将本文提出的模型与时下主流的5组模型进行对比,具体结果如表4所示,其中 WV、PF、PI、POS 分别表示词向量、位置向量、位置信息和词性特征.

表4 综合模型与目前主流模型的对比

模型	特征集	$F1$ (%)
BRNN ^[2]	WV, PI	82.5
BRCNN ^[21]	WV, PI, POS	86.3
BiLSTM-ATT ^[22]	WV, PF	84.0
RCNN-ATT ^[23]	WV, PI, POS	83.7
Hybrid BiLSTM-Siamese ^[24]	WV, PF	81.8
Ensemble	WV, PF	87.7

BRNN: 该方法是 Zhang 等^[2] 在 2015 年使用双向循环神经网络 (BRNN) 进行实验,论文将句子中每个单词的位置信息考虑在内,经过实验得到的 $F1$ 值为 82.5%. 本文使用的模型相较于该模型提高了 5.2%.

BRCNN: 该方法是 Cai 等^[21] 在 2016 年将 CNN 与 RNN 结合起来连接前向和后向 SDP (Shortest Dependency Path, 最短依存路径),同时使用构建细粒度的 Softmax 来提高分类效果,其 $F1$ 达到 86.3%. 本文使用的模型相较于该模型提高了 1.4%.

BiLSTM-ATT: 该方法是 Zhou 等^[22] 在 2016 年,在考虑词嵌入和位置嵌入的基础上,将 Bi-LSTM 与注意力机制结合进行相关实验,其 $F1$ 值得到 84.0%. 本文提出的模型相较于该模型提高了 3.7%.

RCNN-ATT: 该方法是 Zhang 等^[23] 在 2018 年使用 CNN、RNN 和 Attention 进行实验,在考虑词向量

和位置向量的同时,另外加入单词的词性特征信息,其 $F1$ 值达到 83.7%. 本文使用的模型相比于该模型提高了 4.0%.

Hybrid BiLSTM-Siamese: 该方法是 Cui 等^[24] 在 2019 年使用 Siamese 模型的同时,将两个词语层的 BiLSTM 结合在一起,其 $F1$ 值达到 81.80%. 本文使用的模型相比于该模型提高了 5.9%.

4 总结和未来工作

本文提出了一种将两个单一模型 Bi-LSTM 和 CNN 集成到 MLP 的综合模型的方法来提高实体关系抽取性能.该模型不仅可以充分利用两个单一模型的优势,而且还可以再利用 MLP 的自学习能力来提高关系分类的性能.与使用相同数据集中的其他模型相比,本文使用的方法有较高的准确性.

另外,注意力机制可以考虑更多的句子上下文信息,可以更充分的捕获到句子的特征信息,所以下一步计划在本文的模型中添加注意力机制来提高实体关系抽取模型的性能.同时也可以考虑加入一些单词的特征信息,如词性标注、语义角色特征和依存关系等.

参考文献

- Socher R, Huval B, Manning CD, et al. Semantic compositionality through recursive matrix-vector spaces. Proceedings of 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island, Republic of Korea. 2012. 1201–1211.
- Zhang DX, Wang D. Relation classification via recurrent neural network. arXiv: 1508.01006, 2015.
- 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, Ireland. 2014. 2335–2344.
- Miwa M, Bansal M. End-to-end relation extraction using LSTMs on sequences and tree structures. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany. 2016. 1105–1116.
- Yang DD, Wang SZ, Li ZJ. Ensemble neural relation extraction with adaptive boosting. Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden. 2018. 4532–4538.
- Rokach L. Ensemble-based classifiers. Artificial Intelligence

- Review, 2010, 33(1-2): 1-39. [doi: [10.1007/s10462-009-9124-7](https://doi.org/10.1007/s10462-009-9124-7)]
- 7 李冬梅, 张扬, 李东远, 等. 实体关系抽取方法研究综述. 计算机研究与发展, 2020, 57(7): 1424-1448. [doi: [10.7544/issn1000-1239.2020.20190358](https://doi.org/10.7544/issn1000-1239.2020.20190358)]
- 8 宋睿, 陈鑫, 洪宇, 等. 基于卷积循环神经网络的关系抽取. 中文信息学报, 2019, 33(10): 64-72. [doi: [10.3969/j.issn.1003-0077.2019.10.008](https://doi.org/10.3969/j.issn.1003-0077.2019.10.008)]
- 9 Geng ZQ, Chen GF, Han YM, *et al.* Semantic relation extraction using sequential and tree-structured LSTM with attention. Information Sciences, 2020, 509: 183-192. [doi: [10.1016/j.ins.2019.09.006](https://doi.org/10.1016/j.ins.2019.09.006)]
- 10 Chen TT, Wang NB, He M, *et al.* Reducing wrong labels for distantly supervised relation extraction with reinforcement learning. IEEE Access, 2020, 8: 81320-81330. [doi: [10.1109/ACCESS.2020.2990680](https://doi.org/10.1109/ACCESS.2020.2990680)]
- 11 曹春萍, 何亚喆. 融合 BSRU 和 ATT-CNN 的化学物质与疾病的关系抽取方法. 小型微型计算机系统, 2020, 41(4): 794-799. [doi: [10.3969/j.issn.1000-1220.2020.04.022](https://doi.org/10.3969/j.issn.1000-1220.2020.04.022)]
- 12 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. Berlin, Germany. 2016. 2124-2133.
- 13 Liu TY, Wang KX, Chang BB, *et al.* A soft-label method for noise-tolerant distantly supervised relation extraction. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark. 2017. 1790-1795.
- 14 Qin PD, Xu WR, Wang WY. DSGAN: Generative adversarial training for distant supervision relation extraction. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia. 2018. 496-505.
- 15 刘伟, 陈鸿昶, 黄瑞阳. 基于 Tree-based CNN 的关系抽取. 中文信息学报, 2018, 32(11): 34-40. [doi: [10.3969/j.issn.1003-0077.2018.11.005](https://doi.org/10.3969/j.issn.1003-0077.2018.11.005)]
- 16 Gupta P, Rajaram S, Schütze H, *et al.* Neural relation extraction within and across sentence boundaries. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 6513-6520.
- 17 李卫疆, 李涛, 漆芳. 基于多特征自注意力 BLSTM 的中文实体关系抽取. 中文信息学报, 2019, 33(10): 47-56, 72. [doi: [10.3969/j.issn.1003-0077.2019.10.006](https://doi.org/10.3969/j.issn.1003-0077.2019.10.006)]
- 18 Phi VT, Santoso J, Tran VH, *et al.* Distant supervision for relation extraction via piecewise attention and bag-level contextual inference. IEEE Access, 2019, 7: 103570-103582. [doi: [10.1109/ACCESS.2019.2932041](https://doi.org/10.1109/ACCESS.2019.2932041)]
- 19 Peng YF, Rios A, Kavuluru R, *et al.* Chemical-protein relation extraction with ensembles of SVM, CNN, and RNN models. arXiv: 1802.01255, 2017.
- 20 Li DD, Yao SZ, Wang SZ, *et al.* Cross-program design space exploration by ensemble transfer learning. Proceedings of 2017 IEEE/ACM International Conference on Computer-Aided Design. Irvine, CA, USA. 2017. 201-208. [doi: [10.1109/ICCAD.2017.8203779](https://doi.org/10.1109/ICCAD.2017.8203779)]
- 21 Cai R, Zhang XD, Wang HF. Bidirectional recurrent convolutional neural network for relation classification. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany. 2016. 756-765.
- 22 Zhou P, Wei S, Tian J, *et al.* Attention-based bidirectional long short-term memory networks for relation classification. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany. 2016. 207-212.
- 23 Zhang XB, Chen FC, Huang RY. A combination of RNN and CNN for attention-based relation classification. Procedia Computer Science, 2018, 131: 911-917. [doi: [10.1016/j.procs.2018.04.221](https://doi.org/10.1016/j.procs.2018.04.221)]
- 24 Cui ZY, Pan L, Liu SJ. Hybrid BiLSTM-siamese network for relation extraction. Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. Montreal, QC, Canada. 2019. 1907-1909.