

# 基于多粒度和语义信息的中文关系抽取<sup>①</sup>



陈 钰, 张安勤, 许春辉

(上海电力大学 计算机科学与技术学院, 上海 201306)

通讯作者: 张安勤, E-mail: [aqzhang@fudan.edu.cn](mailto:aqzhang@fudan.edu.cn)

**摘 要:** 中文关系抽取采用基于字符或基于词的神经网络, 现有的方法大多存在分词错误和歧义现象, 会不可避免地引入大量冗余和噪音, 从而影响关系抽取的结果. 为了解决这一问题, 本文提出了一种基于多粒度并结合语义信息的中文关系抽取模型. 在该模型中, 我们将词级别的信息合并进入字符级别的信息中, 从而避免句子分割时产生错误; 借助外部的语义信息对多义词进行建模, 来减轻多义词所产生的歧义现象, 并且采用字符级别和句子级别的双重注意力机制. 实验表明, 本文提出的模型能够有效提高中文关系抽取的准确率和召回率, 与其他基线模型相比, 具有更好的优越性和可解释性.

**关键词:** 信息处理; 关系抽取; 注意力机制; 词向量表示; 双向长短期记忆网络

引用格式: 陈钰, 张安勤, 许春辉. 基于多粒度和语义信息的中文关系抽取. 计算机系统应用, 2021, 30(3): 190-195. <http://www.c-s-a.org.cn/1003-3254/7810.html>

## Chinese Relation Extraction Based on Multi-Granularity and Semantic Information

CHEN Yu, ZHANG An-Qin, XU Chun-Hui

(School of Computer Science and Technology, Shanghai University of Electric Power, Shanghai 201306, China)

**Abstract:** Chinese relation extraction adopts character-based or word-based neural networks. Most of the existing methods have word segmentation errors and ambiguity, which will inevitably introduce a lot of redundancy and noise and thus affect the results of relation extraction. In order to solve this problem, this study proposes a Chinese relationship extraction model based on multi-granularity combined with semantic information. In this model, we merge word-level information into character-level information, so as to avoid errors in sentence segmentation; use external semantic information to model polysemous words to reduce the ambiguity caused by semantic words; and adopt Dual attention mechanism at character level and sentence level. The experimental results show that the model proposed in this study can effectively increase the accuracy and recall rate of Chinese relation extraction and has better superiority and interpretability than other baseline models.

**Key words:** information processing; RE; attention; Word2Vec; Bi-LSTM

随着大数据时代的到来, 数据的规模不断增大, 信息过载的问题日益严重. 因此快速准确地抽取关键信息有着重大意义. 关系抽取在信息抽取中有着举足轻重的作用, 目的是提取自然语言句子中实体对之间的语义关系. 实体关系抽取作为自然语言处理的一项基本任务, 是知识图谱、自动问答、机器翻译, 自动文摘

等领域的关键模块. 随着深度学习的不断发展, 引起了人们对 NRE 的兴趣, 现在大多使用神经网络来自动学习语义特征.

### 1 相关工作

作为先驱, Liu 等提了一个基于 CNN 的关系抽取

① 收稿时间: 2020-07-09; 修改时间: 2020-08-11; 采用时间: 2020-08-17; csa 在线出版时间: 2021-03-03

模型<sup>[1]</sup>.在此基础上,Zeng等提出一个带有最大池化层的CNN模型<sup>[2]</sup>,并且引入了位置嵌入来表示位置信息,然后他们设计了PCNN模型<sup>[3]</sup>,但是PCNN模型在句子选择方面存在问题.为了解决这一问题,Lin等<sup>[4]</sup>将注意力机制应用于其中.尽管PCNN模型有着不错的效果,但是它不能像RNN类型的模型一样挖掘上下文的信息.因此,带有注意力机制的LSTM网络也被应用于关系抽取任务中<sup>[5,6]</sup>.

尽管NRE不需要进行特征工程,但是它们忽略了不同语言输入粒度对模型的影响,特别是对于中文关系抽取.根据输入的粒度的不同,现有的中文关系抽取方法可以分为基于字符的关系抽取和基于词的关系抽取的两种.

对于基于字符的关系抽取,它将每个输入语句看作一个字符的序列.这种方法的缺点是不能充分利用词语级别的信息,想比较于基于词语的方法捕获到的特征少.对于基于词语级别的关系抽取,首先要进行分词,导出一个词序列,然后将其输入到神经网络模型中.但是,基于词的模型的性能会受到分割质量的显著影响<sup>[7]</sup>.

比如说,一句中文语句“乔布斯设计部分苹果”有两个实体,“乔布斯”和“苹果”,它们之间的关系为“设计”.在这种情况下,对于这句话的分词为:“乔布斯/设计/部分/苹果”.但是,随着对语句切分的变化,句子的含义可能变得完全不同.如果该句话分割为:“乔布斯设计部/分/苹果”,那么这句话的实体就变为“乔布斯设计部”和“苹果”,它们之间的关系转变为“分发”.因此,无论是基于字符的方法还是基于词语的方法都不能充分利用数据中的语义信息.因此,要从纯文本中发现高层实体关系,需要不同粒度的综合信息的帮助.此外,中文的词语存在大量的多义词,这限制了模型挖掘深层次语义信息的能力.例如,词语“苹果”含有两种不同的含义,即一种水果和电子产品.但是,如果没有语义信息的加入,就很难从纯文本中学习到这种信息.

本文提出了一种能够综合利用句子内部的多粒度信息以及外部知识的网络框架(PL-Lattice)来完成中文关系抽取任务.(1)该模型采用基于Lattice-LSTM模型的结构,将基于词语级别的特征动态的集成到基于字符级别的特征中.因此,可以利用句子中的多粒度信息而不受到分词错误的影响.(2)为了解决中文中多义词的现象,改模型加入了HowNet<sup>[8]</sup>这一外部知识库对词语的多种语义进行标注,在训练阶段结合语义信息,提高模型的挖掘能力.

## 2 问题描述与建模

给定一个中文句子和其中的两个标记实体,中文关系抽取的任务就是提取两个标记实体之间的语义关系.本文提出用于中文关系抽取的PL-Lattice模型,该模型的整体结构如图1所示.

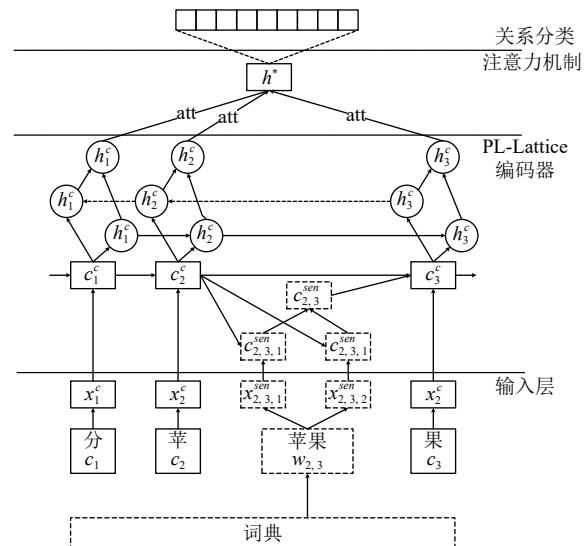


图1 模型整体结构

(1) 输入层: 给定一个以两个目标实体为输入的中文语句,该部分表示语句中的每个单词和字符.该模型可以同时提取和利用字级和词级的信息.

(2) PL-Lattice 编码器: 这一部分使用 Lattice-LSTM 网络结构为基础,改进了字级别和词级别的输入的循环结构,并且将外部语义知识融入到网络里面来实现语义的消歧.

(3) 注意力机制: 使用词级注意力机制和句级注意力机制.

(4) 关系分类层: 通过 *Softmax* 函数输出关系类型.

### 2.1 输入层

本文模型的输入是一个带有两个标记实体的中文句子.为了利用多粒度信息,本文在句子中同时使用字级和词级信息.

#### (1) 字符集向量表示

本文将每个输入的中文语句看作一个字符序列.给定一个由  $M$  个字符组成的句子  $s$ ,表示为  $s = \{c_1, \dots, c_M\}$ ,使用 Skip-gram 模型<sup>[9]</sup>将每个字符映射到一个  $d^c$  维的向量,表示为  $x_i^{ce} \in R^{d^c}$ .

此外,在关系抽取任务中,句子中的字到命名实体

的距离能够影响关系抽取的结果. 所以, 本文采用位置特征来指定句子中的字符, 即当前字符到第一个命名实体和第二个命名实体的相对距离<sup>[2]</sup>. 具体来说, 第  $i$  个字符  $c_i$  到两个命名实体的距离分别表示为  $p_i^1$  和  $p_i^2$ . 对于  $p_i^1$  本文使用以下的方法来计算:

$$p_i^1 = \begin{cases} i - b^1 & i < b^1, \\ 0, & b^1 \leq i \leq e^1 \\ i - e^1, & i > e^1 \end{cases} \quad (1)$$

其中,  $b^1$  和  $e^1$  为第一个命名实体的开始和结束索引,  $p_i^2$  的计算方法与等式 1 类似. 然后,  $p_i^1$  和  $p_i^2$  分别映射为  $d^p$  维度的向量, 表示为  $x_i^{p1} \in R^{d^p}$  以及  $x_i^{p2} \in R^{d^p}$ .

## (2) 词级向量表示

虽然模型有字符特征作为输入, 但是为了充分捕捉句子中的多粒度特征, 本文还需要提取句子中的所有潜在的词级的特征, 潜在词为是由字符组成的句子中的任意子序列. 这些子序列与构建在大型原始文本上的字典  $D$  相匹配得到真正的词级特征.

用  $w_{b,e}$  来表示由第  $b$  个字符开始, 第  $e$  个字符结束的词. 为了将  $w_{b,e}$  表示为向量形式, 大部分文章使用 Word2Vec 模型<sup>[9]</sup> 来将其转化为词向量.

但是, Word2Vec 模型只是将词语映射为一个嵌入向量, 忽略了一词多义这一事实. 本文使用 SAT 模型来解决这一问题, SAT 模型是基于 Skip-gram 模型改进出来的, 它可以同时学习词语以及意义之间的关系, 并将其转换为词向量.

给定一个词语  $w_{b,e}$ , 首先通过检索 HowNet 这一知识库来获得该词语的  $K$  种意义. 用  $Sense(w_{b,e})$  表示词语  $w_{b,e}$  所有意义的集合. 通过 SAT 模型将每种意义  $sen_k^{(w_{b,e})} \in Sense(w_{b,e})$  映射为向量形式, 表示为  $x_{b,e,k}^{sen} \in R^{d^{sen}}$ . 最终,  $w_{b,e}$  表示为一个向量集合, 即  $x_{b,e}^{sen} = \{x_{b,e,1}^{sen}, \dots, x_{b,e,K}^{sen}\}$ .

## 2.2 编码器

本文的编码器是在 Lattice-LSTM 的基础上, 进行改进, 加入词的意义这一特征, 改进了基于字符级别和基于词级别的循环单元的网络结构; 通过改进结构, 减轻了 Lattice-LSTM 编码器中对词的特征的提取导致的字符级别的特征被削弱的现象; 并且使用了双向循环单元的结构, 使编码器能够同时捕捉学习正向和反向的信息, 能够显著提高模型的准确率以及合理性.

### (1) Lattice-LSTM 编码器

LSTM 神经网络是循环神经网络的变种, 主要思

想就是引入一种自适应门控机制来控制 LSTM 单元保留以前状态的同时学习当前数据输入的特征. LSTM 神经网络有 3 个门: 输入门  $i_j$ 、遗忘门  $f_j$  和输出门  $o_j$ . 基于字符的 LSTM 网络表示为:

$$\begin{cases} i_j^c = \sigma(W_i x_j^c + U_i h_{j-1}^c + b_i) \\ o_j^c = \sigma(W_o x_j^c + U_o h_{j-1}^c + b_o) \\ f_j^c = \sigma(W_f x_j^c + U_f h_{j-1}^c + b_f) \\ \tilde{c}_j^c = \tanh(W_c x_j^c + U_c h_{j-1}^c + b_c) \end{cases} \quad (2)$$

$$c_j^c = f_j^c \odot c_{j-1}^c + i_j^c \odot \tilde{c}_j^c \quad (3)$$

$$h_j^c = o_j^c \odot \tanh(c_j^c) \quad (4)$$

其中,  $\sigma()$  为激活函数,  $W$  和  $U$  为可训练权重矩阵,  $b$  为偏置.

在输入语句中给定一个词语  $w_{b,e}$ , 与外部词典  $D$  相匹配, 可以表示为:

$$x_{b,e}^w = e^w(w_{b,e}) \quad (5)$$

其中,  $b$  和  $e$  表示词语在句子中的开始与结束的索引,  $e^w$  表示一个查找表. 在这样的情况下, 对  $c_j^c$  的计算需要结合词语级的表示  $x_{b,e}^w$ , 先构建词语级别的门控逻辑单元, 再与字级的 LSTM 网络相结合, 形成 Lattice-LSTM 编码器. 使用  $c_{b,e}^w$  来表示  $x_{b,e}^w$  的细胞状态.  $c_{b,e}^w$  的计算方法如下:

$$\begin{cases} i_{b,e}^w = \sigma(W_i x_{b,e}^w + U_i h_b^c + b_i) \\ f_{b,e}^w = \sigma(W_f x_{b,e}^w + U_f h_b^c + b_f) \\ \tilde{c}_{b,e}^w = \tanh(W_c x_{b,e}^w + U_c h_b^c + b_c) \end{cases} \quad (6)$$

$$c_{b,e}^w = f_{b,e}^w \odot c_b^c + i_{b,e}^w \odot \tilde{c}_{b,e}^w \quad (7)$$

其中,  $i_{b,e}^w$  和  $f_{b,e}^w$  分别表示词级的门控逻辑单元里面的输入门和遗忘门.

第  $e$  个字符的细胞状态将通过合并以索引  $e$  结束的所有词语的信息来计算, 这些单词  $w_{b,e}$  具有  $b \in \{b' | w_{b',e} \in D\}$  这样的特征. 为了控制每个词语的输入, 设计了一个额外的门  $i_{b,e}^c$ :

$$i_{b,e}^w = \sigma(W x_e^c + U c_{b,e}^w + b^l) \quad (8)$$

第  $e$  个字符的细胞状态计算方法如下:

$$c_e^c = \sum_{b \in \{b' | w_{b',e} \in D\}} \alpha_{b,e}^c \odot c_{b,e}^w + \alpha_e^c \odot \tilde{c}_e^c \quad (9)$$

其中,  $\alpha_{b,e}^c$  以及  $\alpha_e^c$  为归一化因子, 它们的总和等于 1, 计算方法如下:

$$\alpha_{b,e}^c = \frac{\exp(i_{b,e}^c)}{\exp(i_e^c) + \sum_{b' \in \{b' | w_{b',e} \in D\}} \exp(i_{b',e}^c)} \quad (10)$$



$$\alpha_e^c = \frac{\exp(i_e^c)}{\exp(i_e^c) + \sum_{b' \in \{b'' | w_{b'',e} \in D\}} \exp(i_{b',e}^c)} \quad (11)$$

最终,使用式(5)来计算序列中每个字符的最终隐藏状态 $h_j^c$ .这一结构出自文献[7]提出的Lattice LSTM.

## (2) PL-Lattice 编码器

虽然Lattice-LSTM编码器能够利用字符和词的信息,但是它不能充分考虑中文的一词多义的特征.例如,如图1所示, $w_{2,3}$ (苹果)这个词有两种意义, $sen_1^{(w_{2,3})}$ 代表一种水果,但是在Lattice-LSTM中只有一个 $x_{2,3}^w$ 表示 $w_{2,3}$ .并且Lattice-LSTM编码器会造成对词的信息的过度提取会减弱字符级别的信息,甚至会忽略字符的信息,影响模型的效果.

为了解决这两个缺点,本文改进了模型.首先在模型中加入了感知层,将外部语义融入其中,如1.1.2节中所示,本文使用 $x_{b,e,k}^{sen}$ 来词语 $w_{b,e}$ 的第 $k$ 种意义.其次,改进了词级和字符级的LSTM结构,加强了字符之间的信息传递,减弱词级信息的提取.

并且本文采用了双向循环网络的结构,使编码器能同时提取句子上下文的关系与信息对于字符级的循环神经网络单元,本文改进的前向传播的计算方法如下:

$$\begin{cases} i_j^c = \sigma(W_i x_j^c + U_i h_{j-1}^c + V_i C_{j-1}^c + b_i) \\ f_j^c = \sigma(W_f x_j^c + U_f h_{j-1}^c + V_f C_{j-1}^c + b_f) \\ \tilde{c}_j^c = \tanh(W_c x_j^c + U_c h_{j-1}^c + b_c) \end{cases} \quad (12)$$

$$c_j^c = f_j^c \odot c_{j-1}^c + i_j^c \odot \tilde{c}_j^c \quad (13)$$

$$o_j^c = \sigma(W_o x_j^c + U_o h_{j-1}^c + V_o c_j^c + b_o) \quad (14)$$

$$h_j^c = o_j^c \odot \tanh(c_j^c) \quad (15)$$

其中, $\sigma()$ 为激活函数, $W$ 和 $U$ 为可训练权重矩阵, $b$ 为偏置.

这种结构将上一个单元的细胞状态 $c_{j-1}^c$ 合并进入当前单元的各个门控逻辑单元中,增强了 $c_{j-1}^c$ 对 $c_j^c$ 的影响,即增强了字符级别的信息的提取能力.

对于词级的循环神经网络单元,本文的改进的前向传播的计算方法如下:

$$\begin{cases} i_{b,e,k}^{sen} = \sigma(W_i x_{b,e,k}^{sen} + U_i h_b^c + b_i) \\ \tilde{c}_{b,e,k}^{sen} = \tanh(W_c x_{b,e,k}^{sen} + U_c h_b^c + b_c) \end{cases} \quad (16)$$

$$c_{b,e,k}^{sen} = i_{b,e,k}^{sen} \odot c_b^c + (1 - i_{b,e,k}^{sen}) \odot \tilde{c}_{b,e,k}^{sen} \quad (17)$$

其中, $c_{b,e,k}^{sen}$ 表示词语 $w_{b,e}$ 的第 $k$ 个意义的细胞状态.这种结构将遗忘门合并进输入门,会在传入信息的时候

就遗忘一部分信息,从而起到减弱词级信息的作用.

然后,将所有的意义的细胞状态结合起来,得到 $c_{b,e}^{sen}$ ,表示词语 $w_{b,e}$ 所有意义的细胞状态,计算方法如下:

$$c_{b,e}^{sen} = \sum_k \alpha_{b,e,k}^{sen} \odot c_{b,e,k}^{sen} \quad (18)$$

$$\alpha_{b,e,k}^{sen} = \frac{\exp(i_{b,e,k}^{sen})}{\sum_{k'} \exp(i_{b,e,k'}^{sen})} \quad (19)$$

所有感知层的细胞状态都被合并为 $c_{b,e}^{sen}$ ,这样可以更好的表示一个词语多义的现象.然后,类似于式(9)到式(12),以索引 $e$ 结尾的所有词语的细胞状态 $c_{b,e}^{sen}$ 融入到第 $e$ 个字符的细胞状态 $c_e^c$ :

$$c_e^c = \sum_{b \in \{b' | w_{b',e} \in D\}} \alpha_{b,e}^{sen} \odot c_{b,e}^{sen} + \alpha_e^c \odot \tilde{c}_e^c \quad (20)$$

隐含状态 $h$ 的计算方法与式(5)相同.

最终将每一个单元前向传播与后向传播得到的隐含状态结合起来,计算方法如下:

$$h_i = \left[ \vec{h}_i \oplus \overleftarrow{h}_i \right] \quad (21)$$

其中, $\oplus$ 表示将两个向量拼接起来.然后送入注意力层.

## (3) 注意力机制与关系分类

近年来,注意力机制在深度学习的各个领域取得了成功.从本质上讲,深度学习中的注意力机制和人类的选择性注意力类似,都是从众多的信息中选择出对当前任务目标更重要的信息.本文采用了双重注意力机制.

由双向PL-Lattice网络训练产生的输出向量组成矩阵 $h = [h_1, h_2, h_3, \dots, h_M]$ ,其中 $M$ 表示句子的长度.

基于词级的注意力机制的句子表示的计算方法如下所示:

$$H = \tanh(h) \quad (22)$$

$$\alpha = \text{Softmax}(\omega^T H) \quad (23)$$

$$h^* = h \alpha^T \quad (24)$$

其中, $\omega$ 为可训练参数矩阵, $\alpha$ 为 $h$ 所对应的权重向量.

为了计算每种关系相对于句子的权重,使用句子级的注意力机制,将句子 $S$ 的特征向量 $h^*$ 送入Softmax分类器:

$$o = \text{relu}(Wh^* + b) \quad (25)$$

$$P(y|s) = \text{Softmax}(o) \quad (26)$$

其中, $W \in \mathbb{R}^{Y \times d^h}$ 为变换矩阵, $b \in \mathbb{R}^Y$ 为偏执向量. $Y$ 表示

关系类型的数量.  $y$  表示每种类型的概率.

最终, 给定所有的训练样本  $T = (S^{(i)}, y^{(i)})$ , 本文使用交叉熵作为目标函数来计算模型输出结果分别与真实结果分布的差距, 如式 (28) 所示:

$$J(\theta) = \sum_{i=1}^T \log p(y^{(i)} | S^{(i)}, \theta) \quad (27)$$

其中,  $\theta$  表示模型中所有的参数.

### 3 实验数据与参数设定

#### 3.1 实验数据的选取

由于公开的中文关系抽取语料库的匮乏, 目前还没有较为通用且权威的中文远程监督关系抽取数据集.

本文使用的数据为中文的散文数据<sup>[10]</sup>. 该数据集收录 837 篇中文文章, 包含 9 中关系类型, 其中训练集 695 篇, 测试集 84 篇, 验证集 58 篇.

#### 3.2 性能评估指标与参数设定

本文实验采用 3 种评估指标. 召回率,  $F1$  以及  $AUC$ .

召回率 ( $Recall$ ) 是度量的是多个正例被分为正例:

$$Recall = \frac{TP}{TP + FN} \quad (28)$$

式中,  $TP$  表示将正类预测为正类的数量,  $FN$  表示将正类预测为负类的数量.

$F1$  是分类问题的一个衡量指标, 数值在 0~1 之间:

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (29)$$

式中,  $FP$  表示将负类预测为正类的数量.

本文实验参数的设定如表 1 所示.

表 1 实验参数设定

| 参数                      | 数值       |
|-------------------------|----------|
| Learning_rate           | 0.0005   |
| dropout                 | 0.5      |
| char_embedding_size     | 100      |
| lattice_embedding_size  | 200      |
| position_embedding_size | 5        |
| hidden_unit             | 200      |
| epoch                   | 80       |
| regulation              | 1.00e-08 |

### 4 实验结果与分析

为了验证 PL-Lattice 模型在中文实体关系抽取方面的效果, 本文设置了 5 组实验:

(1) BLSTM<sup>[5]</sup>: 提出了一种双向的 LSTM 用于关系抽取.

(2) Att-BLSTM<sup>[6]</sup>: 在双向 LSTM 的基础上加入了词级注意力机制.

(3) PCNN<sup>[3]</sup>: 提出一种具有多实例的分段 CNN 模型.

(4) PCNN+Att<sup>[4]</sup>: 利用注意力机制改进了 PCNN.

(5) Lattice-LSTM<sup>[7]</sup>: 使用基础的 Lattice-LSTM 模型加注意力机制作为对比实验.

实验结果如表 2 所示.

各模型的召回率随训练次数的变化如图 2 所示.

表 2 实验结果

| 模型           | 召回率    | $F1$   | $AUC$  |
|--------------|--------|--------|--------|
| BLSTM        | 0.5178 | 0.6104 | 0.5021 |
| Att-BLSTM    | 0.5273 | 0.5948 | 0.5042 |
| PCNN         | 0.4923 | 0.61   | 0.4826 |
| PCNN+Att     | 0.5232 | 0.6055 | 0.5041 |
| Lattice-LSTM | 0.5822 | 0.6388 | 0.5688 |
| PL-Lattice   | 0.5974 | 0.6757 | 0.5794 |

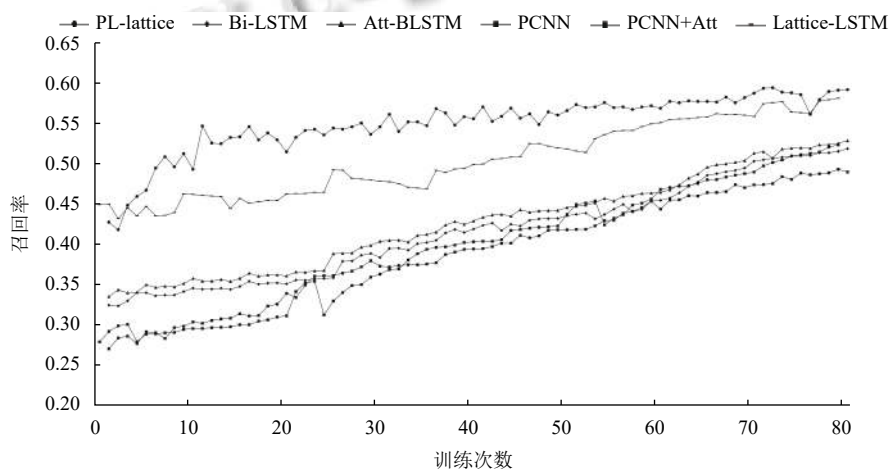


图 2 各模型召回率随训练次数的变化曲线

PL-Lattice 模型的  $F1$  值和 AUC 随训练次数的变化如图 3 所示。

从实验结果可以看出, 注意力机制的加入能够使模型关注到句子中更重要的部分, 从而提升模型的表现能力。由于 LSTM 模型相对于 CNN 模型在处理序列数据上拥有天然的优势, 所以会表现出更好的

水平。

本文提出的 PL-Lattice 模型在各方面都优于其他 5 种模型。经过分析, 认为主要的原因是本文对于分词更加精准, 模型使用了多粒度的信息, 使词向量的表示更加合理化, 并且加入了双重注意力机制, 从多个方面提升了模型的可解释性和能力。

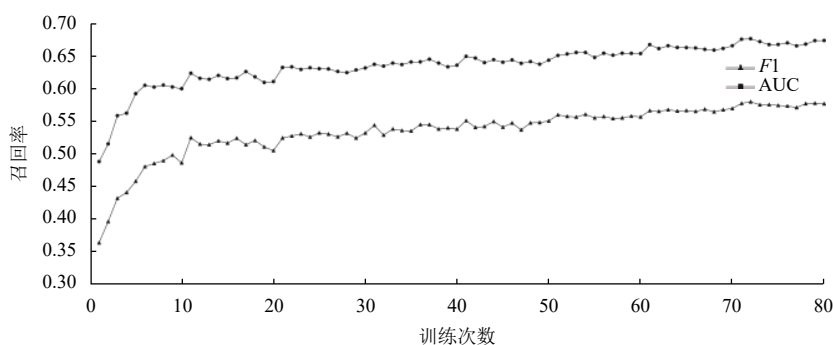


图 3 PL-Lattice 模型  $F1$  和 AUC 随训练次数的变换曲线

## 5 结论与展望

本文提出了一种用于中文关系的 PL-Lattice 模型, 该模型同时使用了字符级别和词级别的信息, 并且引入了外部语义库来表示词向量, 使其拥有更深层次的语义信息, 避免了一词多义的现象。加入了基于词级和基于句子级别的双重注意力机制, 关注了词和句子的多个方面。在散文数据集上与其他 5 种模型进行了对比, 表现出更好的优越性。

在未来, 可以将更多粒度的语料信息融入到模型中, 这些信息可能会由于模型挖掘更深层次的意义特征。

### 参考文献

- Liu CY, Sun WB, Chao WH, *et al.* Convolution neural network for relation extraction. Proceedings of the 9th International Conference on Advanced Data Mining and Applications. Hangzhou, China. 2013. 231–242.
- 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.
- Zeng DJ, Liu K, Chen Y, *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, Portugal. 2015. 1753–1762.
- 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. 2124–2113.
- Zhang DX, Wang D. Relation classification via recurrent neural network. arXiv preprint arXiv: 1508.01006, 2015.
- 刘鉴, 张怡, 张勇. 基于双向 LSTM 和自注意力机制的中文关系抽取研究. 山西大学学报 (自然科学版), 2020, 43(1): 8–13.
- Zhang Y, Yang J. Chinese NER using lattice LSTM. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia. 2018. 1554–1564.
- Qi FC, Yang CH, Liu ZY, *et al.* OpenHowNet: An open sememe-based lexical knowledge base. arXiv preprint arXiv: 1901.09957, 2019.
- Mikolov T, Sutskever I, Chen K, *et al.* Distributed representations of words and phrases and their compositionality. Proceedings of the 26th International Conference on Neural Information Processing Systems. Red Hook, NY, USA. 2013. 3111–3119.
- Xu JJ, Wen J, Sun X, *et al.* A discourse-level named entity recognition and relation extraction dataset for Chinese literature text. arXiv preprint arXiv: 1711.07010, 2019.