

基于双向 LSTM 网络的不确定和否定作用范围识别^{*}

钱忠, 李培峰, 周国栋, 朱巧明



(苏州大学 计算机科学与技术学院, 江苏 苏州 215008)

通讯作者: 钱忠, E-mail: qianzhongqz@163.com; 李培峰, E-mail: pfli@suda.edu.cn

摘要: 不确定和否定信息抽取, 是自然语言处理领域中的重要任务和研究热点. 针对不确定和否定作用范围识别任务, 提出一种基于两层双向 LSTM 神经网络的作用范围识别方法. 首先, 对于从线索词到达词语的句法路径, 使用第 1 层双向 LSTM 神经网络从中学习到有用特征; 接着, 将词法特征与句法路径特征一起组成当前词语的特征表示; 最后, 将作用范围识别问题看作序列标注任务, 利用第 2 层双向 LSTM 神经网络界定当前线索词的作用范围. 实验结果表明, 所提出的模型优于其他神经网络模型, 并在 BioScope 生物医学语料上取得了良好性能. 其中, 在 Abstracts 子语料上的不确定和否定作用范围识别精确率分别达到 86.20% 和 80.28%.

关键词: 信息抽取; 不确定和否定信息; 作用范围识别; LSTM 神经网络; BioScope 语料
中图法分类号: TP18

中文引用格式: 钱忠, 李培峰, 周国栋, 朱巧明. 基于双向 LSTM 网络的不确定和否定作用范围识别. 软件学报, 2018, 29(8): 2427-2447. <http://www.jos.org.cn/1000-9825/5485.htm>

英文引用格式: Qian Z, Li PF, Zhou GD, Zhu QM. Speculation and negation scope detection via bidirectional LSTM neural networks. Ruan Jian Xue Bao/Journal of Software, 2018, 29(8): 2427-2447 (in Chinese). <http://www.jos.org.cn/1000-9825/5485.htm>

Speculation and Negation Scope Detection via Bidirectional LSTM Neural Networks

QIAN Zhong, LI Pei-Feng, ZHOU Guo-Dong, ZHU Qiao-Ming

(School of Computer Science and Technology, Soochow University, Suzhou 215008, China)

Abstract: Speculation and negation information extraction is an important task and research focus in natural language processing (NLP). This paper proposes a two-layer bidirectional long short-term memory (LSTM) neural network model for speculation and negation scope detection. Firstly, a bidirectional LSTM neural network is utilized in the first layer to learn useful feature representations from the syntactic path which is from the cue to the token. Then, lexical features and syntactic path features are concatenated into the feature representations of the token. Finally, taking the scope detection problem as a sequence labeling task, another bidirectional LSTM neural network is employed in the second layer to identify the scope of the current cue. The experimental results show that the presented model is superior to other neural network models and attains excellent performances on BioScope corpus. Particularly, the model achieves the accuracy (percentage of correct scopes) of 86.20% and 80.28% on speculation and negation scope detection on Abstracts subcorpus, respectively.

Key words: information extraction; speculation and negation information; scope detection; LSTM neural network; BioScope corpus

自然语言文本中的信息除了确定的表达以外, 还包含否定、推测等非确定类型的信息. 不确定和否定信息抽取任务即负责从自然语言文本中识别不确定、否定等非事实信息, 以区别事实信息. 其中, 不确定信息的整体

* 基金项目: 国家自然科学基金(61331011, 61472265, 61772354); 江苏省科技计划(BK20151222)

Foundation item: National Natural Science Foundation of China (61331011, 61472265, 61772354); Science and Technology Project of Jiangsu Province (BK20151222)

收稿时间: 2017-03-18; 修改时间: 2017-06-02, 2017-08-27; 采用时间: 2017-11-08; jos 在线出版时间: 2018-01-09

CNKI 网络优先出版: 2018-01-11 17:24:32, <http://kns.cnki.net/kcms/detail/11.2560.TP.20180111.1724.002.html>

或部分具有或然性,在认知程度上介于否定和肯定之间^[1];否定信息包括否定运算符和受其影响且语义反转的文本元素^[2].不确定和否定信息抽取研究是自然语言处理领域很多任务的基础,例如,问答系统^[3,4]、观点检测^[5]、文本蕴含^[6-8]、情感分析^[9]、事件可信度识别^[10,11]等都需要区别事实信息、不确定信息和否定信息.

不确定和否定信息识别主要包含两个子任务.

- 1) 线索词识别(cue detection).线索词是指具有不确定或者否定语义的词语或者短语.例如,例句 1 中的不确定线索词 **may** 和例句 2 中的否定线索词 **not**.本文中,线索词用粗体字表示.
- 2) 作用范围识别(scope detection).作用范围是线索词在语义上管辖或者作用的文本范围,通常情况下是句子中连续的文本片段.例如,例句 1 中不确定线索词 **may** 的作用范围“**may rain any time**”以及例句 2 中否定线索词 **not** 的作用范围“**not cause chronic liver disease**”.本文中,作用范围用下划线表示.

例句 1: It **may rain any time** due to the dark clouds in the sky. (天随时都**可能**下雨,因为乌云已布满天空.)

例句 2: This virus infection does **not cause chronic liver disease** and is rarely fatal. (这种病毒感染**不会引起慢性肝病**且很少致命.)

目前,在线索词识别方面,相关工作已经取得了较好的性能.例如,Tang 等人^[12]使用基于条件随机场(conditional random field,简称 CRF)模型和大规模基于边界(large margin-based model)模型的级联系统,考虑的特征包括词性、命名实体、语义组块等,在 CoNLL2010-Task1B 中,基于 BioScope 语料^[13]的不确定线索词识别任务上获得了最好性能,其 F1 值为 86.36.

传统的不确定和否定作用范围的识别方法主要可分为基于规则的方法^[14,15]、基于统计的方法^[12,16,17]以及规则和统计相结合的方法^[18].其中,在不确定作用范围识别方面,Zou 等人^[16]采用基于树核的方法,在 BioScope 的 Abstracts 子语料上取得的精确率 PCS 为 84.21%,但是在 Full Papers 子语料上的识别精确率 PCS 仅为 67.24%;Velldal 等人^[18]使用了规则和统计结合的模型,在 Full Papers 上获得 PCS 值为 75.15%,然而在 Abstracts 上 PCS 为 79.56%.相对于线索词识别任务,作用范围识别性能有待提高.因此,本文致力于作用范围识别任务.

近年来,神经网络模型已经应用于自然语言处理领域中的各种任务,并且取得了较好的成果,例如语义角色标注^[19,20]、情感分析^[21,22]、事件抽取^[23,24]、关系识别^[25-27]、机器翻译^[28,29]等.其中,基于长短时记忆(long short-term memory,简称 LSTM)单元的神经网络^[30]是循环神经网络(recurrent neural network,简称 RNN)的一种.LSTM 神经网络模型不仅能从句子中学习到有效的特征^[20,21],也可以从句法树路径中获得有用的信息^[25,26].

目前,神经网络模型在不确定和否定作用范围识别任务上的应用还较少^[31,32].本文提出了一种基于两层双向 LSTM 神经网络的不确定和否定作用范围识别方法.首先,对于从线索词到词语的句法路径,使用第 1 层双向 LSTM 神经网络从句法路径中学习到有用特征;接着,将基本词法特征与句法路径特征一起组成当前词语的特征表示;最后,将作用范围识别问题看作序列标注任务,利用第 2 层双向 LSTM 神经网络标注句子中的词语,界定当前线索词的作用范围.实验结果表明,本文提出的模型优于其他神经网络模型,并在 BioScope 生物医学语料上取得了良好性能.

本文第 1 节介绍不确定和否定作用范围以及神经网络模型的相关工作.第 2 节详细阐述本文提出的用于识别不确定和否定作用范围的两层双向 LSTM 神经网络模型.第 3 节介绍 BioScope 语料库和实验设置,并对实验结果进行详细分析.最后,第 4 节给出本文的结论并展望未来的工作.

1 相关工作

本节首先介绍不确定和否定作用范围的相关研究,然后给出神经网络模型的有关成果,最后给出基于神经网络的不确定和否定作用范围识别方法的相关工作.

1.1 传统的不确定和否定作用范围识别方法

早期的不确定和否定作用范围识别工作主要使用规则方法.Chapman 等人^[33]研发了基于正则表达式的系统 NegEx,总结了一张否定线索词表,并认为作用范围由距离线索词 6 个词以内的词语组成.类似的作用范围识别系统还包括健康记录分类系统 NegExpander^[34]、基于电子病历记录的 Maintaining Electronically Health

Records 系统^[35]等.接着出现了基于启发式规则的方法,该方法通过提取线索词和作用范围中词语的实体关系、句法关系、语义关系等信息,总结规则或模板,界定作用范围.Huang 等人^[36]提出了基于句法树的启发式规则模型,判断句法树的词语节点是否属于当前否定线索词的作用范围.Rokach 等人^[37]构建了一种可以自动学习和获取模板的作用范围识别系统.Özgür 等人^[14]使用基于句法特征的启发式规则识别作用范围,在 BioScope 的 Abstracts 和 Full Papers 子语料上达到了 79.98% 和 61.13% 的准确率.Øvreliid 等人^[15]考虑了基于依存句法树结构的启发式规则,在 2010 CoNLL Shared Task 的训练语料(BSP)和验证语料(BSE)上分别获得了 72.31 和 66.73 的 F1 值(线索词已知).基于启发式规则的方法能够取得较好的性能,但是对于具体的任务和不同的语料,需要人工制定模板和规则,鲁棒性(robustness)欠佳.

BioScope 生物医学文献语料库^[13]发布之后,基于机器学习的不确定和否定作用范围识别方法逐渐占据主流地位.Morante 等人^[38]将否定作用范围的识别看作词语分类问题,即判断词语是否位于否定线索词的作用范围之内,主要考虑了词法特征,并提出了一种合适的后处理算法,以确保作用范围是连续的文本.Morante 等人^[39,40]进一步利用浅层句法特征对句中词语进行序列标注,结合了 3 个分类器,并使用投票机制判断词语是否属于不确定^[39]和否定^[40]作用范围.Tang 等人^[12]使用 CRF 模型标注作用范围,考虑的特征包括词语原型、词性、命名实体、语义组块、依存关系等.Zhu 等人^[41]和 Li 等人^[17]认为,线索词的作用范围类似于语义角色标注任务中谓词的覆盖域,可以将线索词看作谓词,将作用范围的识别看作谓词的论元标注问题,提出了一种基于浅层语义分析的作用范围识别模型.Zou 等人^[16]利用基于树核模型提取词语的句法特征,并针对不同词性的线索词构建各自的分类器,在 BioScope 的 Abstracts 子语料上,不确定和否定作用范围识别的精确率分别为 84.21% 和 76.90%.Veldal 等人^[18]提出了一种规则和统计结合的方法,综合考虑了基于依存结构的规则模型和基于 SVM 短语排序器(SVM constituent ranker)的统计模型,在 BioScope 的 Full Papers 子语料上,将不确定和否定作用范围的识别精确率分别提高到 75.15% 和 70.21%.Zou 等人^[42]构建了汉语语料库,方法上,基于序列标注和卷积树核模型,提出了依存特征和依存结构树特征,并用剪枝方法简化结构树,保留有效信息,过滤无用信息,最后用元决策树模型对两模型进行融合.该方法在不确定和否定作用范围识别任务上的精确率分别达到 58.57% 和 68.84%.

1.2 循环神经网络

循环神经网络能够很好地处理序列信息并从中学习有效特征,早期应用于图像^[43]、语音识别等领域^[44].近年来,RNN 在自然语言处理领域的任务上也取得了良好性能.Lin 等人^[45]使用层次结构的 RNN 对文档建模,并提出了一种两步训练方法,在句子调序任务和口语翻译任务上取得了较好的结果.Liu 等人^[46]提出了一种多时间标度(multi time-scale)的 LSTM 神经网络模型提取句子和文档表示,在 4 个数据集上的实验结果表明,该模型优于其他神经网络模型.Zhou 等人^[47]构建了基于注意力(attention)机制的双向 LSTM 网络,从句子中学习特征用于关系分类,仅考虑句子的词向量,而不考虑其他由 NLP 工具产生的词法或者句法特征.在文档级情感分类任务方面,研究人员一般构造层次结构的神经网络模型用于学习文档的特征表示.例如:Tang 等人^[48]先用 LSTM 网络从词向量中学习句子的向量表示,然后用双向的门限 RNN(gated RNN)从句子向量中学习文档的向量表示;Chen 等人^[49]则构造了两层 LSTM 网络,并考虑了用户和产品的注意力机制.

RNN 不仅能够学习到句子中的信息,还可以从句法树路径中学习到有用的特征表示.Roth 等人^[19]利用 LSTM 网络抽取依存路径的特征,用于语义角色标注.在关系分类任务方面:Yu 等人^[25]采用多通道(multichannel)的 LSTM 网络,考虑了实体之间词语、词性、依存关系、上位词(hypernym)这 4 种依存路径;Cai 等人^[26]提出了一种循环-卷积神经网络(recurrent convolutional neural network),该网络中包含有双向的二通道(two-channel) LSTM 网络,学习词语路径、依存关系路径两种依存路径的特征表示.类似地,本文提出的模型中包含一个双向 LSTM 网络用于从句法路径中学习有效特征.

1.3 基于神经网络的不确定和否定作用范围识别方法

近年来,研究人员开始将神经网络模型应用于作用范围识别任务.Fancellu 等人^[31]将作用范围识别问题看作是词序列标注问题,考虑了词语本身、线索词、词性等词法特征,利用双向 LSTM 模型识别作用范围.Qian 等

人^[32]则提出了一种卷积神经网络模型,针对从线索词到当前词的句法路径,在进行卷积操作时考虑了路径中的每个语法单位(包括句中的词语、短语标记、依存关系等)的上下文,提取每个维度的最大值、最小值、加权平均值作为句法路径的特征表示,并结合线索词与当前词位置之差的特征,判断当前词是否位于线索词的作用范围之内.该模型在 BioScope 语料上取得了较好的性能.其中,在 Abstracts 和 Clinical Reports 子语料上,不确定作用范围识别精确率分别为 85.75% 和 73.92%,否定作用范围识别则分别为 77.14% 和 89.66%.

本文是 Qian^[32]工作的扩展,提出了一种两层的双向 LSTM 模型用于识别不确定和否定作用范围.本文方法和 Fancellu 等人^[31]、Qian 等人^[32]、Zou 等人^[42]的工作的主要区别如下.

- 1) Fancellu 等人^[31]基于词法特征,构建了一个基于双向 LSTM 的否定作用范围识别模型.本文在特征和模型结构两方面不同于该系统:首先,除了 Fancellu 等人^[31]使用词语的上下文信息,本文还加入了词语与线索词之间的句法信息,用于获得句子内在的句法结构信息;其次,不同于 Fancellu 等人^[31]使用的单层双向 LSTM 模型仅用于识别否定作用范围,本文提出了一种两层的双向 LSTM 模型用于识别不确定和否定作用范围.该模型先使用一个双向 LSTM 神经网络从句法路径中学习有效的句法路径特征;接着,将词法特征与句法路径特征组成当前词语的特征表示,利用另一个双向 LSTM 神经网络识别当前线索词的作用范围.
- 2) Qian 等人^[32]使用 CNN 从线索词与词语之间的句法路径中学习到了有效信息,从而识别不确定和否定作用范围.本文在采用的特征和神经网络模型上不同于该文:首先,本文在 Qian 等人^[32]使用的句法特征的基础上,把当前词语的上下文信息和句法特征进行和有效的融合,用于识别不确定和否定作用范围;其次,不同于 Qian 等人^[32]使用的 CNN,本文引入了两层的双向 LSTM 模型,实验结果表明,更加有利于作用范围识别任务.
- 3) 语言上,Zou 等人^[42]面向汉语,本文面向英语;模型上,Zou 等人^[42]使用了卷积核模型和元决策树模型,而本文采用了神经网络模型;特征上,与本文类似,Zou 等人^[42]考虑了基于依存句法树的特征.具体来说,Zou 等人^[42]不仅考虑了包括依存句法路径在内的基本依存特征,还从依存句法树中抽取了词原型树、词性树和依存关系树等 3 类特征树结构,并提出了最小完全树、路径树、上下文相关路径树等 3 种剪枝方法简化树结构;而本文从依存句法树中抽取词语到达线索词的最短路径.与 Zou 等人^[42]相比,本文的特征较少且更容易获得.实验结果表明,本文的神经网络模型能够在生物医学语料上取得较好的性能.

2 基于两层双向 LSTM 神经网络的不确定和否定作用范围识别模型

本节详细阐述基于两层双向 LSTM(bidirectional LSTM,简称 BiLSTM)神经网络的不确定和否定作用范围识别模型.首先介绍用于区分词语是否位于作用范围之内的标签框架;接着介绍本文使用的特征,其中,线索词和当前词之间的句法路径特征用一个双向 LSTM 学习得到;最后,使用另一个双向 LSTM 产生句子中各个词语的标签,从而识别当前线索词的作用范围.

2.1 标签框架

本文采用以下 CIO 标签框架.对于每一个词语,为其指定一个标签,用以表示该词是否位于作用范围之内:

- 1) 如果一个词语是线索词(Cue),那么,该词的标签为 C.根据 BioScope 对作用范围的定义,线索词一定位于作用范围之内.因此,本文对线索词单独指定标签,和其他词语加以区分.
- 2) 如果一个词语不是线索词,且位于作用范围之内(In the scope),那么,该词的标签为 I.
- 3) 如果一个词语位于作用范围之外(Out of the scope),那么该词的标签为 O.

例句 3 显示了句中每一个词语的标签(当前线索词是 may):

例句 3: The/O researchers/O claimed/O that/O animals/I may/C be/I affected/I by/I the/I bisphenol/I in/I clinical/I studies/I./O while/O these/O levels/O were/O not/O enough/O to/O cause/O harm/O to/O humans/O ./O

由于线索词一定位于作用范围之内,因此本文使用的标签框架类似于其他工作使用的二类标签框架^[16,31].

一些研究人员使用了 3 类标签框架^[12,39,50],将作用范围的第 1 个单词或者最后一个单词单独予以标记.由于作用范围首尾词与其余单词相比属于绝对少数,因此,这类标签框架有可能导致标签分布不平衡的问题.本文使用的标签框架则可以避免这类问题.

2.2 输入表示

对于每一个词语,本文考虑的特征表示包括与线索词的相对位置特征、线索词特征信息、句法路径特征信息.其中,线索词的相对位置、线索词特征信息为基本特征.句法路径特征由一个双向 LSTM 神经网络学习得到.获得句子中的每一个词语的特征表示之后,整个句子进入另一个双向 LSTM 神经网络,最后交给 softmax 分类器,获得输出结果,即每个词语的每个标签的概率.下面介绍这 3 个特征表示.

- 相对位置(relative position)已被证明在作用范围识别^[32]以及关系分类^[51]、事件抽取^[23]等其他自然语言处理任务中是一个很有用的特征.本文中,每个词语的相对位置被定义为线索词位置与当前词语位置之差.例如,在例句 1 中,线索词 **may** 和当前词语 **any** 的位置之差为-2,如图 1 所示.相对位置特征被映射为一个 d_{rp} 维的向量 f_{rp} .
- 线索词信息用以表示当前词是否为线索词.Fancellu 等人^[31]仅将当前词区分为线索词和非线索词.考虑到每个线索词的作用范围可能有不同的语法结构,我们认为有必要区分不同的线索词.因此,本文的线索词信息定义为:若当前词是线索词,则该特征的取值为词语本身;反之,若当前词不是线索词,则该特征的取值为一个特殊的词语“not_cue”.线索词信息被映射为一个 d_0 维的向量 f_{cue} ,其维度与词向量维度相同.
- 句法路径特征由双向 LSTM 神经网络学习得到.相关工作^[25,26,32]表明,句法路径可以为自然语言处理相关任务提供有效的特征信息.本文的句法路径定义为:在成分句法树或者依存句法树中,从线索词到达当前词的路径.图 1 和图 2 分别显示了例句 1 中从线索词到当前词的成分句法路径、例句 2 中从线索词到当前词的依存句法路径以及线索词与当前词的相对位置、线索词信息等特征.其中,成分句法路径包括词语和成分句法树节点标记,依存句法路径包括词语和依存关系.句法路径可被看成是一个特殊的“句子”, $S=(t_0, t_1, \dots, t_{n-1})$,其中的“词语”可以是原句中的词语、成分句法树节点标记、依存关系.根据一个词向量表 $W_0 \in \mathbb{R}^{d_0 \times |T_0|}$ 将句法路径 S 转化为矩阵 $X_0 \in \mathbb{R}^{d_0 \times n}$.其中, d_0 是词向量的维度, $|T_0|$ 是词向量表的大小(包含的词语个数).下面,句法路径 $X_0 \in \mathbb{R}^{d_0 \times n}$ 进入双向 LSTM 神经网络学习其特征表示.

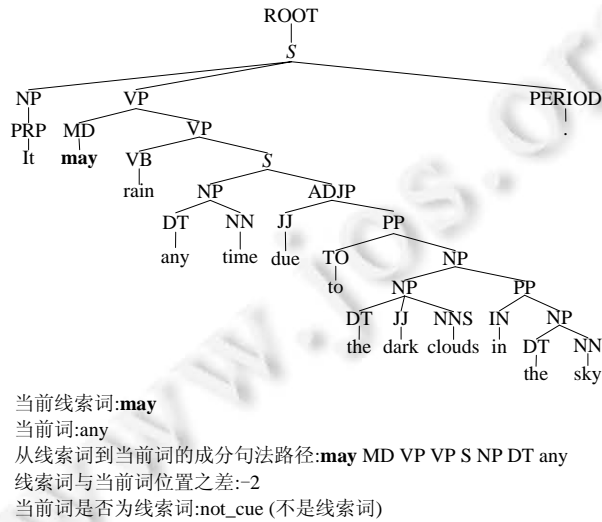


Fig.1 Constituency parse tree of the example sentence 1 and all the features of the current word

图 1 例句 1 的成分句法树和当前词的所有特征

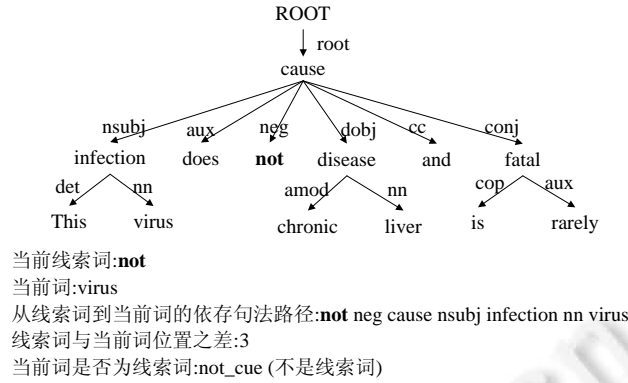


Fig.2 Dependency parse tree of the example sentence 2 and all the features of the current word

图2 例句2的依存句法树和当前词的所有特征

2.3 利用双向LSTM神经网络学习句法路径的特征表示

传统的 RNN 在梯度反向传播阶段存在梯度消失(或称梯度爆炸)和历史信息损失的问题,因此,当传播序列过长、层数过多时,传统 RNN 无法学习到足够的信息.

Hochreiter 等人^[30]提出了 LSTM 单元用以解决传统 RNN 的梯度消失问题.LSTM 专门设计了记忆单元(memory cell)保存历史信息.LSTM 单元包含以下 4 个组成部分:输入门(input gate) i_t 、遗忘门(forget gate) f_t 、输出门(output gate) o_t 和带有自递归连接的神经元 c_t .设 x 为 LSTM 单元的输入, h 为 LSTM 单元隐藏状态的输出.LSTM 单元的状态不仅取决于当前输入 x_t ,还受到上一时刻的输出值 h_{t-1} 的影响.LSTM 单元计算步骤如下:

- 1) 计算输入门的值 i_t 、遗忘门的值 f_t 、输出门的值 o_t .

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

- 2) 计算记忆单元当前状态的候选值 \tilde{c}_t .

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (4)$$

- 3) 根据输入门 i_t 、遗忘门 f_t 、记忆单元上一时刻状态值 c_{t-1} 和当前状态的候选值 \tilde{c}_t ,计算记忆单元当前时刻的状态值.其中, \odot 是对应元素数乘运算.

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (5)$$

- 4) LSTM 单元的输出,即单元的隐藏状态 h_t ,取决于输出门 o_t 和记忆单元当前状态值 c_t :

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

其中,参数 $W_i, W_f, W_o, W_c, U_i, U_f, U_o, U_c \in \mathbb{R}^{d_0 \times d_0}$, $b_i, b_f, b_o, b_c \in \mathbb{R}^{d_0}$, σ 取 logistic sigmoid 函数,LSTM 的隐藏单元的维度为 d_0 .句法路径的词向量矩阵表示 X_0 经过 LSTM 之后,计算得到输出 $H \in \mathbb{R}^{d_0 \times n}$,其中, $H = (h_0, h_1, \dots, h_{n-1})$, $h_j \in \mathbb{R}^{d_0}$ ($0 \leq j \leq n-1$).

然而,标准的 LSTM 以时序顺序处理序列,只考虑了历史信息,忽略了未来信息.许多相关研究^[26,31,47]表明,同时考虑序列的历史信息和未来信息有助于序列建模任务.因此,本文使用双向 LSTM 网络抽取得到正向和反向句法路径的特征表示,其神经网络结构如图 3 所示.双向 LSTM 使用正向、反向两个 LSTM 网络,分别从正反两个方向处理句法路径序列,产生正向句法路径表示 \vec{H}_p 、反向句法路径表示 \overleftarrow{H}_p .句法路径表示 H_p 由 \vec{H}_p 和 \overleftarrow{H}_p 进行对应元素求和运算得到:

$$H_p = \vec{H}_p + \overleftarrow{H}_p \quad (7)$$

为了从 H_p 中获得足够的信息,本文对 H_p 进行最大池化、平均池化和最小池化,即获得 $H_p(r, \cdot)$ 的最大值、

平均值和最小值:

$$h_max(r)=\max[H_p(r,0),H_p(r,1),\dots,H_p(r,n-1)] \quad (8)$$

$$h_avg(r)=\frac{1}{n}\sum_{j=0}^{n-1}H_p(r,j) \quad (9)$$

$$h_min(r)=\min[H_p(r,0),H_p(r,1),\dots,H_p(r,n-1)] \quad (10)$$

其中, $0 \leq r \leq d_0 - 1$. 句法路径的特征表示 h_p 由 h_max, h_avg, h_min 这 3 个特征组成, 并选择 \tanh 作为激活函数:

$$h_p^* = [h_max^T, h_avg^T, h_min^T]^T \quad (11)$$

$$h_p = \tanh(h_p^*) \quad (12)$$

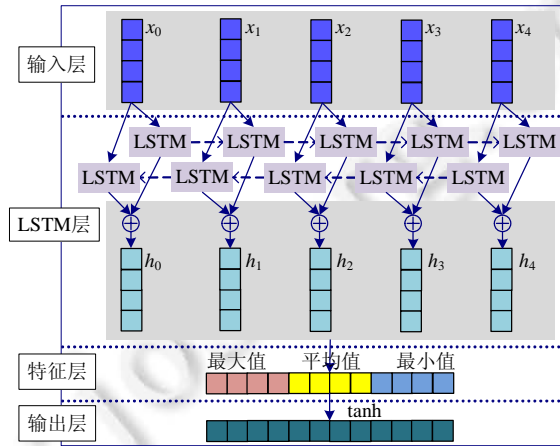


Fig.3 Architecture of the bidirectional LSTM neural network for extracting features from syntactic paths

图 3 用于抽取句法路径特征表示的双向 LSTM 神经网络的结构

2.4 作用范围的识别

本文将作用范围识别问题看作是序列标注问题, 并使用双向 LSTM 网络学习整个句子的特征, 如图 4 所示.

上文已经提到, 对于每一个词语, 本文考虑其 3 个特征: 与线索词的相对位置 f_{rp} 、线索词信息 f_{cue} 和句法路径特征 h_p . 因此, 词语的特征 f_0 可以表示为

$$f_0 = [f_{rp}^T, f_{cue}^T, h_p^T]^T \quad (13)$$

为了减少神经网络模型发生过拟合的可能性, 对于词语的特征表示, 本文考虑了 dropout 操作^[52], 即在模型训练时随机使一定比例的隐藏节点不工作, 其输出置 0:

$$f_1 = f_0 \odot m_0 \quad (14)$$

f_0, f_1 的维度 $\dim(f_0) = \dim(f_1) = d_{rp} + d_0 + 3d_0 = d_{rp} + 4d_0$. $m_0 \in \mathbb{R}^{\dim(f_0)}$ 是一个值由 0 和 1 组成的屏蔽向量(mask vector), 其元素值取 1 的概率为 p , 并遵从伯努利分布(Bernoulli distribution). 已知词语特征表示为 f_1 , 则句子可以表示为 $X_s \in \mathbb{R}^{\dim(f_1) \times n_s}$, n_s 是句子中词语的个数. 为了使句子中的每个词语能够获得来自上下文的信息, 本文再次使用双向 LSTM 网络处理句子, 从而获得句子的正向表示 \vec{H}_s 、反向表示 \overleftarrow{H}_s , 并再次选择 \tanh 作为激活函数, 产生句子的最终表示 H_s^* :

$$H_s^* = \vec{H}_s + \overleftarrow{H}_s \quad (15)$$

$$H_s = \tanh(H_s^*) \quad (16)$$

最后, H_s 进入 softmax 层, 计算得到句子中每个词语的输出, 即每个词语的每个标签(C、I、O)的概率值:

$$O = \text{softmax}(W_s H_s + b_s) \quad (17)$$

其中, $O \in \mathbb{R}^{c \times n_t}$, $W_s \in \mathbb{R}^{c \times \dim(f_1)}$, $b_s \in \mathbb{R}^c$, c 为标签(C、I、O)的类别数目,本文中 $c=3$.

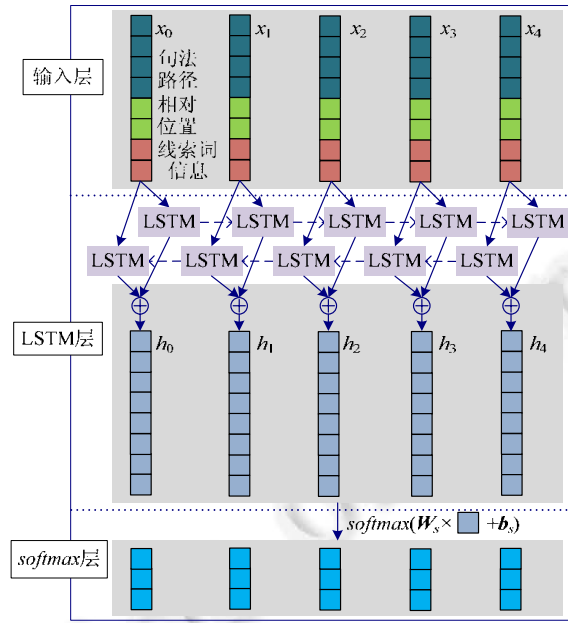


Fig.4 Architecture of bidirectional LSTM neural network for scope detection

图 4 用于识别作用范围的双向 LSTM 神经网络的结构

对于任意向量 a , $softmax$ 函数定义为

$$p(y = i | x, \theta) = \frac{\exp(a_i)}{\sum_{j=0}^{c-1} \exp(a_j)} \quad (18)$$

公式(18)中, $p(y=i|x, \theta)$ 表示样本 x 属于 i 类的概率.

句子中第 t 个词语的向量表示为 $h_t \in \mathbb{R}^{\dim(f_1)}$ ($0 \leq t \leq n_t - 1$), 对应的输出为 $O_t \in \mathbb{R}^c$, 即 O 的第 t 列向量, 其元素可以理解为该词语有可能获得的各个标签(C、I、O)的概率值.

本文采用以下目标函数训练神经网络模型:

$$J(\theta) = -\frac{1}{m} \sum_{i=0}^{m-1} \log p(y_i | x_i, \theta) + \frac{\lambda}{2} \|\theta\|^2 \quad (19)$$

其中, $p(y_i|x_i, \theta)$ 表示训练样本 x_i 的标注标签 y_i 的概率, m 为训练样本个数, λ 为正则化系数, θ 为模型中所有需要学习的参数的集合.

根据 BioScope 的定义, 作用范围是连续的文本. 而本文将作用范围识别任务看作是词语的序列标注问题, 则可能导致识别得到若干不连续的作用范围片段. 例如, 在 BioScope 的 Abstracts 子语料上的实验结果表明, 当使用成分句法路径时, 不确定和否定作用范围识别结果中, 不连续的作用范围分别占 5.30% 和 5.47%. 因此, 与 Zou 等人^[16]、Qian 等人^[32]类似, 本文采用 Morante 等人^[38]提出的后处理算法, 以确保所识别的作用范围的连续性.

3 实验

本节首先介绍 BioScope 语料库的概况, 然后给出实验设置, 包括衡量性能的指标、使用的工具、神经网络模型中超参数的设置等. 接着介绍基准系统, 用于和本文的两层双向 LSTM 模型进行对比实验. 最后, 给出本文提出的模型在 BioScope 语料库上的性能和分析, 并与前人的一系列先进系统进行比较.

3.1 BioScope语料库

本文选择 BioScope 生物医学语料库^[13]验证文中提出的基于两层双向 LSTM 网络的作用范围识别方法.该语料库标注了不确定和否定线索词及其作用范围.为了较为全面地反映生物医学领域的语言现象,BioScope 包含了以下 3 个子语料库:生物医学领域的论文摘要(以下简称 Abstracts 子语料)、临床医学领域的放射性报告(Clinical Radiology Reports,以下简称 Clinical 子语料)、生物医学领域的论文全文(以下简称 Full Papers 子语料).其中,Full Papers 子语料包含 9 篇文章,5 篇来自 FlyBase 数据库,4 篇来自公开的 BMC Bioinformatics 期刊论文网站.Abstracts 来自 GENIA 语料库.科技文献摘要各种文本挖掘任务的主要目标语料.因此,Abstracts 子语料是 BioScope 中规模最大的子语料库.

表 1 显示了 BioScope 的相关统计数据,其中,文本长度以 1 个单词为 1 个单位.Abstracts 和 Full Papers 子语料均来自于生物医学领域的论文,在句子结构和语言风格上类似,所以这两个子语料库的句子长度和作用范围的长度也相近.分不同类型的作用范围来看,Abstracts 和 Full Papers 的否定作用范围的平均长度(分别为 8.60 和 8.35)均小于不确定作用范围(分别为 16.10 和 14.38).Clinical 语料来自于临床诊断,句式相对随意,句子和作用范围长度比 Abstracts 和 Full Papers 都要短.Clinical 子语料中不确定句子和作用范围的平均长度分别为 11.96 和 5.92,而 Abstracts 分别为 29.77 和 16.10,Full Papers 则分别为 30.76 和 14.38.Clinical 子语料的否定句子和作用范围的平均长度分别为 8.53 和 4.87,而 Abstracts 分别为 29.28 和 8.60,Full Papers 语料则分别为 30.55 和 8.35.此外,从表 1 中可知,BioScope 语料中,线索词与作用范围左边界距离远远小于其与右边界的距离,说明大多数线索词靠近作用范围的左端.

Table 1 Statistics of BioScope corpus

表 1 BioScope 语料库的统计数据

		Abstracts	Clinical records	Full papers
总计	文档数	1 273	1 954	9
	句子数	11 871	6 383	2 670
	句子平均长度	25.47	7.71	24.54
不确定	句子数	2 101	855	519
	作用范围数	2 659	1 112	672
	句子平均长度	29.77	11.96	30.76
	作用范围平均长度	16.10	5.92	14.38
	线索词与作用范围左边界平均距离	1.06	0.82	1.66
	线索词与作用范围右边界的平均距离	14.04	4.10	11.71
否定	句子数	1 597	865	339
	作用范围数	1 719	870	376
	句子平均长度	29.28	8.53	30.55
	作用范围平均长度	8.60	4.87	8.35
	线索词与作用范围左边界平均距离	0.97	0.10	1.08
	线索词与作用范围右边界的平均距离	6.63	3.77	6.27

3.2 实验设置和基准系统

为了能够与已有工作进行公平的比较,本文在 Abstracts 子语料上进行 10 倍交叉验证,而 Clinical 和 Full Papers 子语料上的结果由 Abstracts 上训练得到的模型产生,从而检验本文提出的模型在生物医学领域内不同文本类型之间的鲁棒性.因此,Abstracts 子语料上的结果是域内(in-domain)检验,而 Clinical 和 Full Papers 子语料的结果是跨域(cross-domain)检验.

传统的准确率(precision)、召回率(recall)和 $F1$ 值($F1$ -measure)用于反映作用范围识别任务中词语的性能.为了能够更加准确地描述作用范围识别性能,本文采用被正确识别的作用范围的百分比(percentage of correct scopes,简称 PCS)作为衡量指标.此外,本文还采用被正确识别的作用范围左边界百分比(percentage of correct left boundaries,简称 PCLB)和右边界百分比(percentage of correct right boundaries,简称 PCRB),用于进一步描述作用范围识别性能.

本文使用斯坦福句法分析工具(Stanford Parser, <http://nlp.stanford.edu/software/lex-parser.shtml>)生成成分句法树和依存句法树,并使用 GENIA Treebank 1.0 语料库^[53]作为训练语料生成句法分析模型,在 GENIA Treebank 1.0 上,10 倍交叉验证的结果为 $F1=87.12$. GENIA Treebank 1.0 语料库对 GENIA 语料库进行了基于 Penn Treebank II 格式的短语结构的句法标注.而 BioScope 中规模最大的 Abstracts 子语料来自于 GENIA 语料.因此,本文使用的句法分析模型在 BioScope 上会取得令人满意的性能.

对于本文提出的两层双向 LSTM 模型,其中的超参数设置为 $d_0=100, d_{pp}=10$. 本文选择带冲量(momentum)的随机梯度下降(stochastic gradient descent)算法训练神经网络模型.

为了比较不同模型的性能,本文采用以下模型作为基准系统(baseline).

- 1) 基于规则的模型(简称 Rules):该规则模型基于句子的成分句法树.设 P 为当前线索词的祖父节点(即短语标记节点, chunk node),若 P 管辖的叶子节点只有当前线索词,则将 P 的父节点作为新的 P ;否则,作用范围是 P 管辖的叶子节点的集合.
- 2) 最大熵分类模型(简称 MaxEnt):采用 Zou 等人^[16]使用的平面句法特征(flat syntactic features)判断词语是否属于当前线索词的作用范围,其中包括 6 个基本特征、10 个成分句法特征、5 个依存句法特征.本文使用的最大熵分类器工具是 Mallet(<http://mallet.cs.umass.edu/>).
- 3) 简单的卷积神经网络模型(简称 CNN+CNN):该模型使用 CNN 抽取句法路径特征表示,其卷积层单元数与词向量维度相同,即取 $d_0=100$;同时,采用另一个 CNN 隐藏层学习句子中每个词的特征表示,其输入单元数与每个词的特征表示向量的维度相同,取 $\dim(f_0)=d_{pp}+d_0+3d_0=d_{pp}+4d_0=410$.
- 4) 单向 LSTM(unidirectional LSTM)神经网络模型(简称 UniLSTM+UniLSTM):该神经网络结构与本文使用的双向神经网络模型基本相同,唯一的区别在于 LSTM 网络是单向的.

类似地,本文的双向 LSTM 神经网络模型则记为 BiLSTM+BiLSTM.

3.3 实验性能和分析(Abstracts子语料)

表 2 显示了本文提出的模型在 Abstracts 子语料上的性能,其中, Cpath 表示采用的结构化特征是成分句法树中从线索词到词语的最短路径,即成分句法路径; Dpath 表示采用的结构化特征是依存句法树中从线索词到词语的最短路径,即依存句法路径; Cpath+Dpath 表示采用的结构化特征为成分句法路径+依存句法路径.

Table 2 Performances of various models on Abstracts subcorpus

表 2 Abstracts 子语料上各种模型的性能

作用范围类型	系统	P (%)	R (%)	$F1$	PCLB (%)	PCRB (%)	PCS (%)	
不确定	Rules	93.55	88.16	90.76	73.60	85.26	67.96	
	MaxEnt	94.71	90.54	92.56	84.81	85.11	72.47	
	CNN+CNN	Cpath	91.34	94.06	92.65	82.25	83.12	69.84
		Dpath	92.29	90.94	91.60	85.00	78.90	68.82
		Cpath+Dpath	93.01	92.49	92.73	83.75	79.69	68.52
	UniLSTM+UniLSTM	Cpath	95.74	96.01	95.87	89.06	89.55	80.67
		Dpath	94.87	94.24	94.54	91.05	82.55	76.24
		Cpath+Dpath	95.81	94.74	95.25	90.07	86.76	79.43
	BiLSTM+BiLSTM	Cpath	95.68	96.85	96.26	93.61	90.38	86.20
		Dpath	94.80	94.73	94.73	90.60	86.84	79.54
		Cpath+Dpath	95.15	95.91	95.52	91.05	88.08	81.46
	否定	Rules	74.28	63.27	68.21	33.10	51.13	18.44
		MaxEnt	85.46	72.95	78.63	84.00	58.29	46.42
		CNN+CNN	Cpath	86.48	82.22	84.09	78.48	73.77
Dpath			87.08	85.67	86.23	87.78	74.58	65.62
Cpath+Dpath			87.62	86.96	87.11	86.74	76.79	66.78
UniLSTM+UniLSTM		Cpath	92.39	84.84	88.15	81.27	86.21	69.11
		Dpath	91.49	89.66	90.39	91.10	82.55	74.99
		Cpath+Dpath	92.20	89.92	90.96	90.98	84.00	76.32
BiLSTM+BiLSTM		Cpath	94.04	91.66	92.71	93.02	86.45	80.11
		Dpath	92.59	91.12	91.80	94.51	86.45	80.28
		Cpath+Dpath	93.77	86.98	89.87	90.40	83.07	75.28

从表 2 中可以看出:

- 1) 在 Abstracts 上统计模型(包括最大熵模型和神经网络模型)取得的不确定作用范围识别性能高于否定作用范围识别性能.在句子长度接近的情况下(Abstracts 子语料中,包含不确定信息的句子和包含否定信息的句子,其平均长度分别为 29.77 和 29.28),否定作用范围的平均长度(8.60)远小于不确定作用范围的平均长度(16.10).因此,相对于不确定作用范围,属于否定作用范围的词语较少,即正例较少.词语的正负例不平衡导致否定作用范围识别的性能低于不确定作用范围.
- 2) 在不确定作用范围识别任务方面,简单的规则模型所达到的性能接近于最大熵模型和 CNN+CNN 模型.这是由于许多不确定线索词(例如 suggest、may、indicate 等,这 3 个词在 Abstracts 子语料中占不确定线索词的 53.70%)用于表达观点.因此,这些不确定线索词的作用范围位于该线索词所在的短语结构之内,易用规则模型识别.而该规则模型在否定作用范围识别任务上的性能很低($PCS=18.44\%$),且其 $PCLB(33.10\%)$ 远远低于 $PCRB(51.13\%)$,即作用范围左边界识别性能较差,说明否定作用范围不能简单地用线索词所在的短语结构进行识别.而最大熵模型识别否定作用范围的性能也低于神经网络模型,主要是最大熵模型不能很好地识别否定作用范围的右边界,其 $PCRB(58.29\%)$ 远低于 $PCLB(84.00\%)$.
- 3) 相对于传统的规则模型和最大熵模型,本文的 BiLSTM+BiLSTM 模型对性能的提升较为明显.其中,在否定作用范围识别任务上, PCS 值的提升分别为 61.84% 和 33.86%;而在不确定作用范围识别任务上, PCS 值的提升分别为 18.24% 和 13.73%.由此可见,BiLSTM+BiLSTM 模型更有助于提升否定作用范围识别性能.特别地,相对于最大熵模型,BiLSTM+BiLSTM 模型有效地提高了否定作用范围的 $PCRB$ 值,在使用成分句法路径和依存句法路径时, $PCRB$ 均提高了 28.16%.
- 4) 在不确定和否定作用范围识别任务上,BiLSTM+BiLSTM 模型均明显优于 CNN+CNN 模型和 UniLSTM+UniLSTM 模型.这说明相对于不考虑路径中词语上下文的 CNN 模型和只考虑上文的 UniLSTM 模型,BiLSTM 同时考虑了词语的上下文环境.因此,BiLSTM 可以从句法路径和句子中学习更加有效的特征表示.
- 5) BiLSTM+BiLSTM 模型同时考虑成分句法路径和依存句法路径并不能提高作用范围识别的性能.我们认为,这是由于 BiLSTM 已经能够从成分句法路径或者依存句法路径中学习足够的特征信息.同时从成分句法路径和依存句法路径学习特征信息反而可能会造成过拟合.

为了进一步探究 Abstracts 子语料上不确定和否定作用范围识别任务的性能,下面,本文列举 Abstracts 中出现次数占据前 10 位的不确定和否定线索词的作用范围识别性能,并进行分析.

图 5 显示了本文的神经网络模型在使用成分句法路径时,Abstracts 中出现次数前 10 位的不确定线索词的作用范围识别性能(图中,每个线索词的作用范围识别性能有 3 个,其对应的模型从左到右分别为 CNN+CNN, UniLSTM+UniLSTM, BiLSTM+BiLSTM,下同;对于 Abstracts 子语料中的各线索词,本文取其 10 倍交叉验证的平均 PCS 值,下同).出现次数最多的 3 个线索词,即 suggest, may 和 indicate, 占据了 Abstracts 中不确定线索词的 53.70%.其中, suggest 和 indicate 通常用于表达一定的观点或者阐述某些情况, may、might、whether、possible 则通常用于引导具有不确定语义的子句或者短语.因此,这些线索词的作用范围往往是具有完整语义的文本片段.在成分句法树中,这些线索词的作用范围一般位于该词的短语句法标记节点管辖的子树之内.因此,在使用成分句法路径时,这些线索词的作用范围识别性能较好.

然而, or 的 PCS 值远低于其他不确定线索词.其中, CNN+CNN 模型在 or 的作用范围识别方面的性能较差, PCS 值小于 10%.我们认为,这是由于 or 的运用比较灵活,可以用于连接两个概念相近的语义成分,例如,连接名词、动词、形容词、短语或者是子句.因此,判断 or 的作用范围较困难.此外, CNN+CNN 模型对 appear, could, likely 的作用范围识别性能也低于其他高频线索词,在使用成分句法路径时, PCS 值分别只有 39.95%, 44.63%, 51.55%, 且其 $PCLB$ 值(分别为 58.90%, 57.26%, 56.13%)远低于 $PCRB$ 值(分别为 75.02%, 80.52%, 93.42%). UniLSTM+UniLSTM 对于 could, likely 的作用范围性能的提升也不是很明显.这些线索词的 PCS 值高低取决于其 $PCLB$ 值,

识别其作用范围的左边界是一个难点.经过对语料的分析,我们认为,是由于 Abstracts 子语料中,appear,could,likely 等词用法灵活,存在于作用范围的不同位置,如下列例句所示.其中,尖括号“<>”代表 BiLSTM+BiLSTM 模型识别的作用范围.例句 4 中,appear 位于作用范围的左边界并引导从句;而例句 5 中,appear 出现在 appear to be done 结构中,其作用范围包含了该子句的主语.

例句 4: It also <appears that NF-kappaB suppression by RCC-S may contribute to the immunosuppression of host immunity>.

例句 5: We furthermore demonstrate that this process involves activation of the transcription factor NF-chi B and, to a lesser extent, of NF-IL6, while <the activity of the transcription factor AP-1, <shown to otherwise confer IL-6 inducibility, appeared to be unaffected by LTB4>.

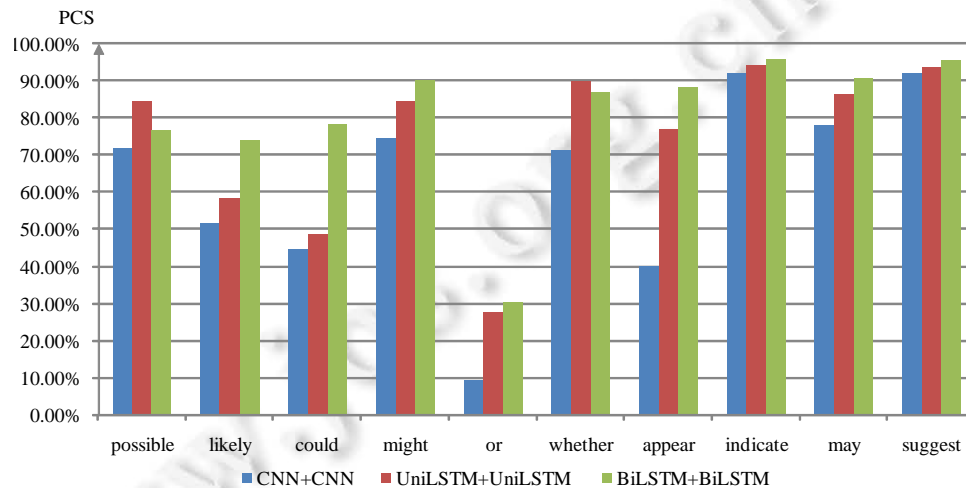


Fig.5 PCS of scopes of the most frequent ten speculative cues on Abstracts subcorpus (with constituency syntactic path/Cpath)

图 5 Abstracts 子语料中频数前 10 位的不确定线索词的作用范围 PCS 值(使用成分句法路径)

本文的 BiLSTM+BiLSTM 模型对 or 的 PCS 值提高明显,但是仍然远低于其他不确定线索词.而 BiLSTM+BiLSTM 模型将 appear,could,likely 的 PCS 值分别提高到 88.13%,78.22%,73.85%,其 PCLB 值则分别提高到 92.40%,84.52%,82.59%,说明 BiLSTM+BiLSTM 模型能够很好地识别位于这些线索词左边的词语是否在其作用范围之内,即能有效地界定其作用范围的左边界.

图 6 显示了本文的神经网络模型在使用依存句法路径时,Abstracts 子语料中出现次数前 10 位的否定线索词的作用范围识别性能.其中,not 和 no 占据了否定线索词的 73.07%.因此,not 和 no 的作用范围识别性能在很大程度上决定了 Abstracts 上否定作用范围的性能.一方面,不少否定线索词的用法相对固定.例如,CNN+CNN 模型已经可以在识别 without 的作用范围上获得较高的性能,BiLSTM+BiLSTM 模型对其提升不是很明显.这是因为 without 后面通常接名词或者名词性短语,其作用范围的识别相对容易.另一方面,否定线索词通常用于否定短语或者长度较短的子句.相对于不确定线索词,否定线索词的作用范围长度较短(Abstracts 子语料上否定作用范围平均长度为 8.60,not 和 no 的作用范围平均长度分别为 8.31 和 9.31),其句法结构相对简单,且作用范围之内的词语与线索词之间存在着较紧密的依存关系.所以,大部分否定线索词的作用范围识别取得了较好的性能.

然而,否定线索词 unable 和 cannot 的作用范围识别性能较低,BiLSTM+BiLSTM 模型取得的 PCS 值分别为 23.33%和 55.56%,其 PCLB 值则分别为 38.33%和 63.89%,远低于其 PCRB 值 76.67%和 91.67%.与上文分析的不确定线索词 appear,could,likely 类似,正确识别 unable 和 cannot 的作用范围,重点在于正确判断这两个否定线索词左边的文本片段是否属于其作用范围.

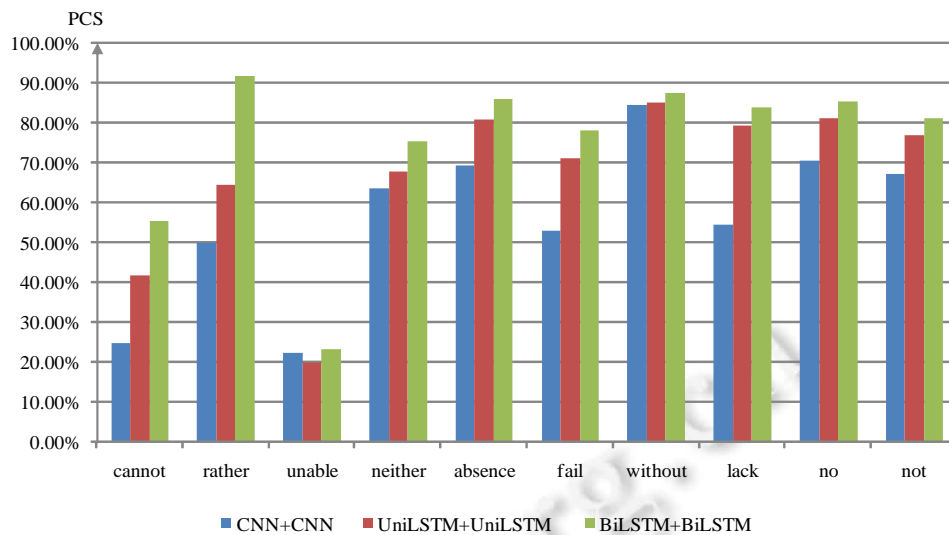


Fig.6 PCS of scopes of the most frequent ten negative cues on Abstracts subcorpus (with dependency syntactic path/Dpath)

图 6 Abstracts 子语料中频数前 10 位的否定线索词的作用范围 PCS 值(使用依存句法路径)

3.4 实验性能和分析(Clinical子语料)

根据上文的分析,Abstracts 和 Full Papers 子语料分别来自生物医学领域论文的摘要和全文部分,因此,这两个子语料的句子和作用范围都较长,而 Clinical 子语料的语言风格区别于 Abstracts 和 Full Papers,其句子和作用范围均较短.表 3 展示了 Clinical 上作用范围识别任务的性能.

Table 3 Performances of various models on Clinical subcorpus

表 3 Clinical 子语料上各种模型的性能

作用范围类型	系统	P (%)	R (%)	F1	PCLB (%)	PCRB (%)	PCS (%)	
不确定	Rules	85.93	83.07	84.47	73.20	64.66	50.90	
	MaxEnt	91.87	81.06	86.13	78.15	73.74	56.56	
	CNN+CNN	Cpath	86.49	87.90	87.19	72.30	80.13	59.26
		Dpath	85.12	87.28	86.18	78.51	75.63	61.33
		Cpath+Dpath	84.93	88.33	86.60	75.27	78.96	61.87
	UniLSTM+UniLSTM	Cpath	84.39	92.04	88.05	77.25	83.63	65.38
		Dpath	88.27	90.62	89.43	83.00	78.60	68.26
		Cpath+Dpath	89.56	88.16	88.85	74.37	80.40	62.41
	BiLSTM+BiLSTM	Cpath	89.53	92.69	91.80	83.27	86.06	73.11
		Dpath	87.33	89.09	88.20	76.71	76.98	62.68
		Cpath+Dpath	84.88	91.11	87.88	72.39	80.58	60.25
	否定	Rules	95.19	61.22	74.52	92.87	44.60	39.54
MaxEnt		98.01	76.63	86.01	97.59	54.60	53.10	
CNN+CNN		Cpath	93.78	95.07	94.42	96.44	87.59	85.06
		Dpath	92.48	92.28	92.38	95.17	86.78	82.53
		Cpath+Dpath	91.36	93.79	92.56	95.52	86.44	82.76
UniLSTM+UniLSTM		Cpath	96.60	93.75	95.15	97.24	88.85	86.21
		Dpath	96.32	95.63	95.97	98.62	90.11	88.97
		Cpath+Dpath	96.56	94.12	95.33	98.51	89.31	88.05
BiLSTM+BiLSTM		Cpath	96.16	96.34	96.25	97.59	90.80	88.74
		Dpath	95.05	97.99	96.50	97.59	94.14	92.30
		Cpath+Dpath	95.83	97.69	96.75	98.16	94.02	92.07

上文已经提到,以 Abstracts 为训练集生成模型,并将 Clinical 作为测试集,得到 Clinical 上的作用范围识别结果.实验结果同样表明,BiLSTM+BiLSTM 在 PCS 值上优于其他方法.

本文的 BiLSTM+BiLSTM 在使用成分句法路径时,在 Clinical 子语料上不确定作用范围的 PCS 值达到最好的 73.11%,远远低于其在 Abstracts 上的性能($PCS=86.20\%$).为了进一步分析实验结果,与 Abstracts 子语料类似,本文探索了 Clinical 子语料上若干高频线索词的作用范围性能.

图 7 左图显示了在使用成分句法路径的情况下,Clinical 子语料中出现频数前 5 位的不确定线索词的作用范围识别性能.这 5 个词共占据总的不确定线索词的 52.88%.可以看到,其不确定线索词分布与 Abstracts 子语料大不相同:Clinical 子语料中出现最频繁的不确定线索词为 or,占有所有不确定线索词的 23.29%;BiLSTM+BiLSTM 模型取得 PCS 值只有 57.14%,远低于大部分其他不确定线索词(may、evaluate、consistent、suggest 等).这是由于在作为训练集的 Abstracts 子语料中,or 仅占不确定线索词的 3.42%,模型很难从训练集中学习到关于 or 的作用范围的足够的特征信息.此外,正如上文的分析,or 的运用比较灵活,可以连接各类语法元素.

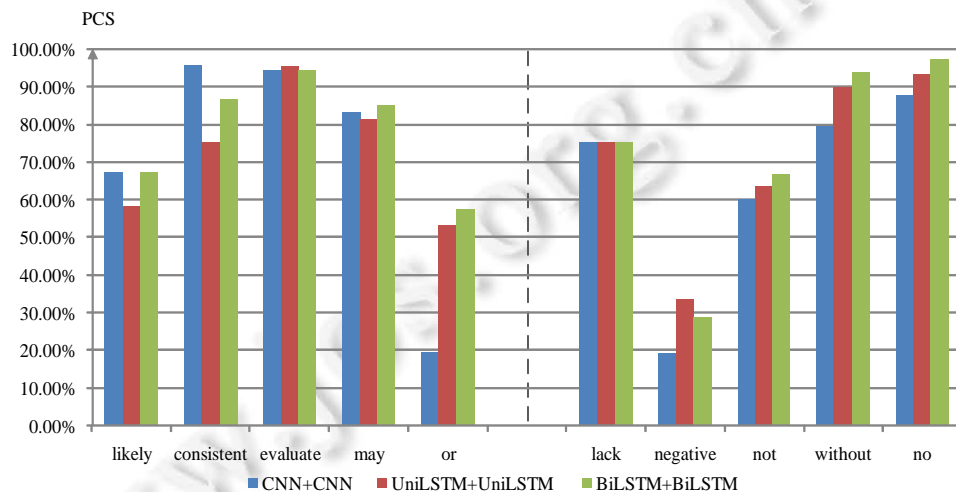


Fig.7 PCS of scopes of the most frequent 5 speculative (left part) and negative (right part) cues on Clinical subcorpus (with constituency syntactic path/Cpath and dependency syntactic path/Dpath, respectively)

图 7 Clinical 子语料中频数前 5 位的不确定(左)和否定(右)线索词的作用范围 PCS 值 (分别使用成分句法路径和依存句法路径)

此外,图 7 右图显示了在使用依存句法路径的情况下,Clinical 子语料中出现频数前 5 位的否定线索词其作用范围的性能.Clinical 中的否定线索词中 no 最多,占据了 77.59%,not、without、no、negative 是出现频度前 4 位的否定线索词,占 98.05%,其余否定线索词的出现频度均小于 5 次.由于 Clinical 中否定线索词分布类似于作为训练语料的 Abstracts 子语料,主要的高频否定线索词 no、without、not 在 Abstracts 中也属于出现次数前 5 位的否定线索词,且 Clinical 中句子和作用范围均较短,句法结构并不复杂,因此,本文的 BiLSTM+BiLSTM 模型在识别 Clinical 子语料的否定作用范围任务上表现得很好,其 PCS 值为 92.30%.

3.5 实验性能和分析(Full Papers子语料)

表 4 是 Full Papers 子语料的作用范围识别性能.实验结果表明,BiLSTM+BiLSTM 取得的 $F1$ 和 PCS 值优于本文的其他方法,除了在否定作用范围识别任务上 $F1$ 值(86.06)略低于 UniLSTM+UniLSTM(86.16).其中,使用成分句法路径时,本文的 BiLSTM+BiLSTM 模型在该子语料的不确定作用范围识别任务上达到最好的 $PCS=63.10\%$;使用依存句法路径时,BiLSTM+BiLSTM 模型在否定作用范围识别任务上达到最好的 $PCS=62.50\%$.相对于 Abstracts 和 Clinical 子语料,Full Papers 子语料上的不确定作用范围和否定作用范围性能均较低.Full Papers 子语料中含有不确定和否定信息的句子长度分别为 30.76 和 30.55,均大于 Abstracts(分别为 29.77,29.28)和 Clinical 子语料(分别为 11.96,8.53).因此,Full Papers 子语料中的句子具有较为复杂的句法结构,识别句子中线

索词的作用范围难度更大。

从表 4 中可知,在不确定作用范围识别任务方面,BiLSTM+BiLSTM 模型取得的 $PCLB=86.76\%$,远远高于其 $PCRB=70.83\%$;而在否定作用范围识别任务方面,BiLSTM+BiLSTM 模型达到的 $PCLB=87.23\%$,也高于其 $PCRB=71.28\%$ 。因此,Full Papers 子语料上的作用范围识别任务性能较低,主要是由于本文的模型在识别作用范围右边界方面表现得不太好。主要原因是 Full Papers 子语料的句子中,不少句子末尾有带括号的词或词组,代表图例和引用文献。根据 BioScope 所采取的标注规则,这类带括号的词或词组不包含在作用范围之内。本文 BiLSTM+BiLSTM 模型错误地将部分句子末尾带括号的词识别为作用范围的一部分。在不确定和否定作用范围识别任务上,这类错误分别占 13.73% 和 12.50%。

Table 4 Performances of various models on Full Papers subcorpus

表 4 Full Papers 子语料上各种模型的性能

作用范围类型	系统	P (%)	R (%)	$F1$	$PCLB$ (%)	$PCRB$ (%)	PCS (%)	
不确定	Rules	84.71	73.07	78.46	58.18	60.42	39.58	
	MaxEnt	87.49	75.76	81.21	71.43	60.12	41.96	
	CNN+CNN	Cpath	86.17	83.13	84.62	69.05	61.01	43.45
		Dpath	84.23	78.82	81.43	75.45	56.25	43.45
		Cpath+Dpath	84.94	82.72	83.81	75.30	58.93	45.54
	UniLSTM+UniLSTM	Cpath	87.50	87.46	87.48	74.85	66.52	50.60
		Dpath	90.73	70.10	79.09	77.68	50.89	37.95
		Cpath+Dpath	89.77	82.68	86.08	81.70	62.20	51.19
	BiLSTM+BiLSTM	Cpath	90.39	86.65	88.48	86.76	70.83	63.10
		Dpath	88.62	82.40	85.40	80.21	61.16	49.85
		Cpath+Dpath	84.71	89.42	87.00	81.10	65.63	54.02
	否定	Rules	68.82	64.70	66.70	43.09	48.94	18.88
		MaxEnt	83.19	60.53	70.07	77.66	45.21	31.65
		CNN+CNN	Cpath	76.70	78.15	77.42	68.62	54.26
Dpath			78.56	78.69	78.62	76.86	56.12	41.76
Cpath+Dpath			79.73	80.06	79.89	76.33	61.44	46.28
UniLSTM+UniLSTM		Cpath	87.16	80.25	83.56	76.60	68.35	50.53
		Dpath	87.98	84.42	86.16	83.51	66.76	55.32
		Cpath+Dpath	87.62	84.52	86.04	85.37	66.76	57.45
BiLSTM+BiLSTM		Cpath	91.38	79.36	84.94	86.97	65.16	55.05
		Dpath	87.72	84.45	86.06	87.23	71.28	62.50
		Cpath+Dpath	86.50	85.35	85.92	87.23	65.96	56.65

3.6 实验性能和分析(输入)

表 5 反映了不同输入对于作用范围识别性能的影响。

第 2.2 节中已经提到,本文将特征分为基本特征和句法路径特征。其中,基本特征包括了词语与线索词的相对位置、当前词是否为线索词。在不确定作用范围识别任务上,相对于其他句法路径,使用成分句法路径 Cpath 能够在各子语料上取得最好的性能,因此,表 5 中本文选择成分句法路径。类似地,在否定作用范围识别任务上,使用依存句法路径 Dpath 能够在各子语料上取得最好的性能,表 5 中选择依存句法路径。

表 5 显示,如果仅考虑基本特征而忽略句法路径,作用范围识别性能远远低于考虑句法路径的情况。这说明作用范围识别性能主要取决于词语与线索词之间的句法信息,不同线索词的作用范围,其句法结构也不同。如果仅考虑句法路径特征,性能略低于同时考虑基本特征和句法路径,但高于仅考虑基本特征。这表明基本特征有助于提升作用范围识别性能。基本特征可以将线索词与非线索词加以区分,并描述词语与线索词之间的距离。不同线索词的作用范围平均长度也不同。例如,在 Abstracts 子语料的不确定线索词中,用于表达观点和论述事实的 suggest,经常用于引导从句,其作用范围平均长度为 19.62;而通常用于修饰名词性词组或者短语的 possible,其作用范围平均长度仅有 12.96。

Table 5 Performances of various types of input on BioScope**表 5** BioScope 上不同输入的性能

作用范围类型	子语料	特征	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i>	PCLB (%)	PCRB (%)	PCS (%)
不确定	Abstracts	基本	87.29	91.63	89.36	80.48	80.78	67.36
		Cpath	95.26	96.13	95.68	90.22	89.02	81.65
		基本+Cpath	95.68	96.85	96.26	93.61	90.38	86.20
	Clinical Reports	基本	79.47	92.80	85.62	68.44	78.51	55.85
		Cpath	86.93	91.82	89.31	76.98	83.18	66.10
		基本+Cpath	89.53	92.69	91.80	83.27	86.06	73.11
Full Papers	基本	85.12	80.08	82.52	72.62	58.18	43.01	
	Cpath	89.87	88.05	88.95	82.29	66.96	56.55	
	基本+Cpath	90.39	86.65	88.48	86.76	70.83	63.10	
否定	Abstracts	基本	84.11	73.11	78.02	84.24	56.55	46.42
		Dpath	93.51	89.48	91.37	91.16	85.57	78.13
		基本+Dpath	92.59	91.12	91.80	94.51	86.45	80.28
	Clinical Reports	基本	92.27	98.32	95.20	96.44	93.22	89.89
		Dpath	94.80	97.14	95.96	97.36	93.45	91.49
		基本+Dpath	95.05	97.99	96.50	97.59	94.14	92.30
Full Papers	基本	72.93	70.21	71.55	78.19	35.64	27.39	
	Dpath	89.67	80.22	84.68	83.24	66.49	53.46	
	基本+Dpath	87.72	84.45	86.06	87.23	71.28	62.50	

3.7 与领域内先进系统的比较

表 6 显示了本文的 BiLSTM+BiLSTM 模型与近几年领域内性能最优的若干作用范围识别系统在 BioScope 上性能的比较.其中,由于 Velldal 等人^[18]在其基于 SVM 短语排序器的统计模型的基础上引入了大量规则,为了公平比较,本文仅例举他们的统计模型的性能.另外,Fancellu 等人^[31]使用的基于词法特征的双向 BiLSTM 模型,针对的是柯南道尔的夏洛克·福尔摩斯(Conan Doyle's Sherlock Holmes)语料.本文重新实现了他们的系统,并应用到 BioScope 语料库上.

Table 6 Performance comparison of our system with the state-of-the-art ones in PCS (%) on BioScope**表 6** 本文系统与领域内先进系统在 BioScope 上的 PCS 值(%)的性能比较

	系统	Abstracts	Clinical reports	Full papers
不确定	Velldal 等人 ^[18] (2012)	75.48	73.61	66.17
	Zou 等人 ^[16] (2013)	84.21	72.92	67.24
	Fancellu 等人 ^[31] (2016)	71.27	56.47	48.51
	Qian 等人 ^[32] (2016)	85.75	73.92	59.82
	本文 BiLSTM+BiLSTM	86.20	73.11	63.10
否定	Velldal 等人 ^[18] (2012)	68.35	89.59	60.90
	Zou 等人 ^[16] (2013)	76.90	85.31	61.19
	Fancellu 等人 ^[31] (2016)	69.05	90.80	49.20
	Qian 等人 ^[32] (2016)	77.14	89.66	55.32
	本文 BiLSTM+BiLSTM	80.28	92.30	62.50

可以看到,本文的 BiLSTM+BiLSTM 模型在 BioScope 否定作用范围识别任务上均高于其他系统,说明本文模型更适合识别 BioScope 上的否定作用范围.对于不确定作用范围,在 BioScope 的 Abstracts 子语料上,本文取得了最好的性能,即 PCS=86.20%,比 Qian 等人^[32]高出 0.45%.

在 Clinical 子语料的不确定作用范围识别任务上,本文模型达到 PCS=73.11%,比 Qian 等人^[32]的 PCS=73.92%低了 0.81%.Velldal 等人^[18]的统计模型取得的 PCS 为 73.61%,如果结合规则模型,则 PCS 可以达到 78.69%.如第 3.5 节中的分析,主要原因是 BiLSTM+BiLSTM 模型在识别线索词 or 的作用范围上表现得不好.而在否定作用范围识别方面,本文的 PCS 值为 92.30%,优于其他所有模型.即使是 Velldal 等人^[18]的统计与规则结合的模型,其 PCS 值也只达到 90.74%,其简单规则模型却可以达到 91.43%,但均低于本文模型.

本文在 Full Papers 子语料上取得的不确定和否定作用范围的 PCS 性能分别为 63.10%和 62.50%,远远低于 Abstracts 和 Clinical 子语料.其他系统所取得的 PCS 也均低于 70%.说明识别 Full Papers 子语料上的作用范围是一个相对困难的任务.Zou 等人^[16]和 Veldal 等人^[18]均考虑了以当前词语为根的子树结构,在该子语料上的不确定作用范围识别性能方面优于本文模型.

为了进一步验证本文提出的神经网络模型的通用性,本文选择 CoNLL-2010 Shared Task^[54]中任务 2 所使用的 Evaluation 语料库.与 Full Papers 类似,该语料库同样来自生物医学领域论文,其中仅标注了不确定信息.根据 CoNLL-2010 Shared Task 的规定,Evaluation 语料的结果由 Abstracts 和 Full Papers 上训练得到的模型产生.此外,本文将 Qian 等人^[32]的系统应用于该语料,获得其不确定作用范围识别性能;与表 6 类似,为了与本文模型进行公平比较,本文例举了 Veldal 等人^[18]的统计模型的性能.

表 7 显示了基于 Evaluation 语料的不确定作用范围识别任务上本文模型和领域内先进系统性能的比较.在选择成分句法路径 Cpath 时,本文的神经网络模型能够取得最好的性能.Evaluation 语料上含有不确定信息的句子和作用范围长度分别为 34.41 和 16.75,均大于 Full Papers 子语料(分别为 30.76 和 14.38).因此,与 Full Papers 类似,Evaluation 语料上句子较长,句法结构复杂,作用范围也较难识别.本文模型在 Evaluation 语料上的不确定作用范围识别的 PCS 值为 65.34%,优于领域内其他系统.Veldal 等人^[18]的统计模型获得的 PCS 为 58.37%,若结合规则模型,则 PCS 可以达到 69.60%.可见,相对于 Abstracts 和 Clinical 子语料,各模型在 Evaluation 上的性能也很低.类似于 Full Papers,Evaluation 中也含有很多带括号的名词性词组或短语,表示引用文献、图例、公式、缩写等.这些词组既可能出现在句子末尾,也可能出现在句子中间,导致句子结构更加复杂.根据标注规则,这些词组或者位于作用范围之内,或者不属于作用范围.而本文模型得到的错例中,将带括号词组中部分词语识别为属于作用范围,这类错误占 22.35%.

Table 7 Performance comparison of our system with the state-of-the-art ones in PCS (%) on Evaluation corpus

表 7 本文系统与领域内先进系统在 Evaluation 语料上的 PCS 值(%)的性能比较

系统	PCS (%)	
领域内先进系统	Veldal 等人 ^[18] (2012)	58.37
	Fancellu 等人 ^[31] (2016)	49.95
	Qian 等人 ^[32] (2016)	55.95
本文系统	Rules	50.63
	MaxEnt	51.89
	CNN+CNN	51.11
	UniLSTM+UniLSTM	61.37
	BiLSTM+BiLSTM	65.34

4 结论与未来的工作

本文针对自然语言文本中线索词的作用范围识别这一自然语言处理领域内的重要任务,提出了一种基于两层双向 LSTM 神经网络的模型.首先,本文利用一个双向 LSTM 神经网络模型从线索词与词语之间的句法路径中学习到有效的特征表示;然后,综合考虑句法路径特征和其他词法特征,组成各个词语的特征表示;最后,将作用范围识别问题转化为序列标注问题,利用另一个双向 LSTM 神经网络的模型处理整个句子,标注各个词语,识别当前线索词的作用范围,并采用适当的后处理算法确保作用范围是连续的文本.与其他基于神经网络模型的作用范围识别方法(Fancellu 等人^[31]和 Qian 等人^[32])相比,本文的模型既考虑了线索词与词语之间的句法结构特征信息,又考虑了词语的上下文信息.实验结果表明,本文提出的基于两层双向 LSTM 神经网络的模型在 BioScope 语料上能够取得良好的性能.其中,在不确定作用范围识别方面,本文在 Abstracts 和 Clinical 子语料上能取得的 PCS 值分别为 86.20%和 73.11%;在否定作用范围识别方面,本文模型优于领域内所有其他先进系统,Abstracts 和 Clinical 上的 PCS 值则分别为 80.28%和 92.30%.在句子结构较为复杂的 Full Papers 子语料上,本文的模型也能取得良好的性能,其不确定和否定作用范围 PCS 值分别为 63.10%和 62.50%.在语言风格类似 Full Papers 的 Evaluation 子语料上,本文模型取得的 PCS 值分别为 65.34%,高于领域内其他先进系统.本文还将双向

LSTM 模型与 CNN 模型、单向 LSTM 模型进行了实验对比.由于双向 LSTM 模型同时考虑了序列的上下文,能够从序列中学习到的有效信息,因此实验结果显示,双向 LSTM 模型优于 CNN 模型和单向 LSTM 模型.

然而,本文仍然使用基于单词的识别模型.其他基于短语结构的模型(例如 Veldal 等人^[18])在识别作用范围时同样能够取得较好的性能.此外,本文的神经网络模型在生物医学领域内跨不同风格文本的实验中,其鲁棒性不足,在 Full Papers 子语料上取得的性能与前人的最优系统类似,效果提升不明显.因此,未来本文将尝试构建基于短语的作用范围识别系统,以识别结构复杂、语义完整的句子中的作用范围.本文还将构建更加有效的神经网络模型,提高作用范围识别系统的性能.

References:

- [1] Lakoff G. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 1973,2(4): 458–508. [doi: 10.1007/BF00262952]
- [2] Morante R, Sporleder C. Modality and negation: An introduction to the special issue. *Computational Linguistics*, 2012,38(2): 223–260. [doi: 10.1162/COLI_a_00095]
- [3] Pustejovsky J, Knippen B, Littman J, Saurí R. Temporal and event information in natural language text. *Language Resources and Evaluation*, 2005,39(2):123–164. [doi: 10.1145/1314493.1314495]
- [4] Angeli G, Nayak N, Manning CD. Combining natural logic and shallow reasoning for question answering. In: Erk K, Smith NA, eds. *Proc. of the 2016 Conf. on the 54th Annual Meeting of the Association for Computational Linguistics (Vol.1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 2016. 442–452.
- [5] Wiebe J, Wilson T, Cardie C. Annotating expressions of opinions and emotions in language. *Language Resources & Evaluation*, 2005,39(2-3):165–210. [doi: 10.1007/s10579-005-7880-9]
- [6] de Marneffe MC, Maccartney B, Grenager T, Cer D, Rafferty A, Manning CD. Learning to distinguish valid textual entailments. In *Proc. of the 2nd Pascal Challenges Workshop on Recognising Textual Entailment (RTE-2)*. 2006.
- [7] Hickl A, Bensley J. A discourse commitment-based framework for recognizing textual entailment. In: Dagan I, Dolan B, Giampiccolo D, Magnini B, eds. *Proc. of the Workshop on Textual Entailment and Paraphrasing*. Stroudsburg: Association for Computational Linguistics, 2007. 171–176.
- [8] Saurí R, Pustejovsky J. Determining modality and factuality for text entailment. In: *Proc. of the 1st IEEE Int'l Conf. on Semantic Computing (ICSC 2007)*. 2007. 509–516. [doi: 10.1109/ICSC.2007.80]
- [9] Councill IG, McDonald R, Velikovich L. What's great and what's not: Learning to classify the scope of negation for improved sentiment analysis. In: Morante R, Sporleder C, eds. *Proc. of the Workshop on Negation and Speculation in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2010. 51–59.
- [10] Saurí R. A factuality profiler for eventualities in text. [Ph.D. Thesis]. Waltham: Brandeis University, 2008.
- [11] Saurí R, Pustejovsky J. Are you sure that this happened? Assessing the factuality degree of events in text. *Computational Linguistics*, 2012,38(2):261–299. [doi: 10.1162/COLI_a_00096]
- [12] Tang BZ, Wang XL, Wang X, Yuan B, Fan FX. A cascade method for detecting hedges and their scope in natural language text. In: Farkas R, Vincze V, Szarvas G, Móra G, Csirik J, eds. *Proc. of the 14th Conf. on Computational Natural Language Learning-Shared Task*. Stroudsburg: Association for Computational Linguistics, 2010. 13–17.
- [13] Vincze V, Szarvas G, Farkas R, Móra G, Csirik J. The BioScope corpus: Biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 2008,9(Suppl.11):1–9. [doi: 10.1186/1471-2105-9-S11-S9]
- [14] Özgür A, Radev DR. Detecting speculations and their scopes in scientific text. In: Koehn P, Mihalcea R, eds. *Proc. of the 2009 Conf. on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2009. 1398–1407.
- [15] Øvrelid L, Veldal E, Oepen S. Syntactic scope resolution in uncertainty analysis. In: Huang CR, Jurafsky D, eds. *Proc. of the 23rd Int'l Conf. on Computational Linguistics*. Stroudsburg: Association for Computational Linguistics, 2010. 1379–1387.
- [16] Zou BW, Zhou GD, Zhu QM. Tree kernel-based negation and speculation scope detection with structured syntactic parse features. In: Baldwin T, Korhonen A, eds. *Proc. of the 2013 Conf. on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2013. 968–976.

- [17] Li JH, Zhou GD, Wang HL, Zhu QM. Learning the scope of negation via shallow semantic parsing. In: Huang CR, Jurafsky D, eds. Proc. of the 23rd Int'l Conf. on Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2010. 671–679.
- [18] Vellidal E, Øvrelid L, Read J, Oepen S. Speculation and negation: Rules, rankers, and the role of syntax. *Computational Linguistics*, 2010,38(2):369–410. [doi: 10.1162/COLI_a_00126]
- [19] Roth M, Lapata M. Neural semantic role labeling with dependency path embeddings. In: Erk K, Smith NA, eds. Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Vol.1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2016. 1192–1202.
- [20] Zhou J, Xu W. End-to-End learning of semantic role labeling using recurrent neural networks. In: Zong CQ, Strube M, eds. Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing (Vol.1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2015. 1127–1137.
- [21] Xu JC, Chen DL, Qiu XP, Huang XJ. Cached long short-term memory neural networks for document-level sentiment classification. In: Duh K, Carreras X, eds. Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2016. 1660–1669.
- [22] Wang XY, Jiang WJ, Luo ZY. Combination of convolutional and recurrent neural network for sentiment analysis of short texts. In: Matsumoto Y, Prasad R, eds. Proc. of the 26th Int'l Conf. on Computational Linguistics: Technical Papers (COLING 2016). Stroudsburg: Association for Computational Linguistics, 2016. 2428–2437.
- [23] Chen YB, Xu LH, Liu K, Zeng DJ, Zhao J. Event extraction via dynamic multi-pooling convolutional neural networks. In: Zong CQ, Strube M, eds. Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing (Vol.1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2015. 167–176.
- [24] Nguyen TH, Cho K, Grishman R. Joint event extraction via recurrent neural networks. In: Nenkova A, Rambow O, eds. Proc. of the 2016 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 2016. 300–309.
- [25] Xu Y, Mou LL, Li G, Chen YC, Peng H, Jin Z. Classifying relations via long short term memory networks along shortest dependency paths. In: Callison-Burch C, Su J, eds. Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2015. 1785–1794.
- [26] Cai R, Zhang XD, Wang HF. Bidirectional recurrent convolutional neural network for relation classification. In: Erk K, Smith NA, eds. Proc. of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2016. 756–765.
- [27] Yang YL, Tong YH, Ma SL, Deng ZH. A position encoding convolutional neural network based on dependency tree for relation classification. In: Duh K, Carreras X, eds. Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2016. 65–74.
- [28] Zhou J, Cao Y, Wang XG, Li P, Xu W. Deep recurrent models with fast-forward connections for neural machine translation. *Trans. of the Association for Computational Linguistics*, 2016,4:371–383.
- [29] Wang MX, Lu ZD, Li H, Liu Q. Memory-Enhanced decoder for neural machine translation. In: Duh K, Carreras X, eds. Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2016. 278–286.
- [30] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997,9(8):1735–1780. [doi: 10.1162/neco.1997.9.8.1735]
- [31] Fancellu F, Lopez A, Webber B. Neural networks for negation scope detection. In: Erk K, Smith NA, eds. Proc. of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2016. 495–504.
- [32] Qian Z, Li PF, Zhu QM, Zhou GD, Luo ZC, Luo W. Speculation and negation scope detection via convolutional neural networks. In: Duh K, Carreras X, eds. Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2016. 815–825.
- [33] Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, 2001,34(5):301–310. [doi: 10.1006/jbin.2001.1029]

- [34] Aronow D, Feng FF, Croft WB. Research paper: Ad hoc classification of radiology reports. *Journal of the American Medical Informatics Association*, 1999,6(5):393–411. [doi: 10.1136/jamia.1999.0060393]
- [35] Boytcheva S, Strupchanska A, Paskaleva E, Tcharaktchiev D. Some aspects of negation processing in electronic health records. In: Paskaleva E, Piperidis S, eds. *Proc. of the Int'l Workshop Language and Speech Infrastructure for Information Access in the Balkan Countries Held in Conjunction with RANLP 2005*. INCOMA, 2005. 1–8.
- [36] Huang Y, Lowe HJ. Research paper: A novel hybrid approach to automated negation detection in clinical radiology reports. *Journal of the American Medical Informatics Association*, 2007,14(3):304–311. [doi: 10.1197/jamia.M2284]
- [37] Rokach L, Romano R, Maimon O. Negation recognition in medical narrative reports. *Information Retrieval Journal*, 2008,11(6): 499–538. [doi: 10.1007/s10791-008-9061-0]
- [38] Morante R, Liekens A, Daelemans W. Learning the scope of negation in biomedical texts. In: Lapata M, Ng HT, eds. *Proc. of the 2008 Conf. on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2008. 715–724.
- [39] Morante R, Daelemans W. Learning the scope of hedge cues in biomedical texts. In *Proc. of the BioNLP 2009 Workshop*. Stroudsburg: Association for Computational Linguistics, 2009. 28–36.
- [40] Morante R, Daelemans W. A metalearning approach to processing the scope of negation. In *Proc. of the 13th Conf. on Computational Natural Language Learning (CoNLL 2009)*. Stroudsburg: Association for Computational Linguistics, 2009. 21–29.
- [41] Zhu QM, Li JH, Wang HL, Zhou GD. A unified framework for scope learning via simplified shallow semantic parsing. In: Li H, Márquez L, eds. *Proc. of the 2010 Conf. on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2010. 714–724.
- [42] Zou BW, Qian Z, Chen ZC, Zhu QM, Zhou GD. Negation and uncertainty information extraction oriented to natural language text. *Ruan Jian Xue Bao/Journal of Software*, 2016,27(2):309–328 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4860.htm> [doi: 10.13328/j.cnki.jos.004860]
- [43] Cierniak R. A new approach to image reconstruction from projections using a recurrent neural network. *Applied Mathematics and Computer Science*, 2008,18(2):147–157. [doi: 10.2478/v10006-008-0014-y]
- [44] Juang CF, Chiou CT, Huang HJ. Noisy speech recognition by hierarchical recurrent neural fuzzy networks. In: *Proc. of Int'l Symp. on Circuits and Systems (ISCAS 2005)*. IEEE, 2005. 5122–5125. [doi: <http://dx.doi.org/10.1109/ISCAS.2005.1465787>]
- [45] Lin R, Liu SJ, Yang MY, Li M, Zhou M, Li S. Hierarchical recurrent neural network for document modeling. In: Callison-Burch C, Su J, eds. *Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2015. 899–907.
- [46] Liu PF, Qiu XP, Chen XC, Wu SY, Huang XJ. Multi-Timescale long short-term memory neural network for modelling sentences and documents. In: Callison-Burch C, Su J, eds. *Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2015. 2326–2335.
- [47] Zhou P, Shi W, Tian J, Qi ZY, Li BC, Hao HW, Xu B. Attention-Based bidirectional long short-term memory networks for relation classification. In: Erk K, Smith NA, eds. *Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Vol.2: Short Papers)*. Stroudsburg: Association for Computational Linguistics, 2016. 207–212.
- [48] Tang DY, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification. In: Callison-Burch C, Su J, eds. *Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2015. 1422–1432.
- [49] Chen HM, Sun MS, Tu CC, Lin YK, Liu ZY. Neural sentiment classification with user and product attention. In: Carreras X, Duh K, eds. *Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2016. 1650–1659.
- [50] Lapponi E, Velldal E, Øvrelid L, Read J. UiO2: Sequence-labeling negation using dependency features. In: Agirre E, Bos J, Diab M, eds. *Proc. of the 1st Joint Conf. on Lexical and Computational Semantics—Vol.1: Proc. of the Main Conf. and the Shared Task, and Vol.2: Proc. of the 6th Int'l Workshop on Semantic Evaluation*. Stroudsburg: Association for Computational Linguistics, 2012. 319–327.

- [51] Zeng DJ, Liu K, Lai SW, Zhou GY, Zhao J. Relation classification via convolutional deep neural network. In: Tsujii J, Hajic J, eds. Proc. of the 25th Int'l Conf. on Computational Linguistics: Technical Papers (COLING 2014). Stroudsburg: Association for Computational Linguistics, 2014. 2335–2344.
- [52] Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov R. Improving neural networks by preventing co-adaptation of feature detectors. CoRR, abs/1207.0580, 2012.
- [53] Tateisi Y, Yakushiji A, Ohta T, Tsujii J. Syntax annotation for the GENIA corpus. In: Dale R, Wong KF, Su J, Kwong YO, eds. Proc. of 2nd Int'l Joint Conf. on Natural Language Processing (IJCNLP 2005). Stroudsburg: Association for Computational Linguistics, 2005. 220–225.
- [54] Farkas R, Vincze V, Móra G, Csirik J, Szarvas G. The conll-2010 shared task: learning to detect hedges and their scope in natural language text. In Farkas R, Vincze V, Szarvas G, Móra G, Csirik J, eds. Proc. of the 14th Conf. on Computational Natural Language Learning-Shared Task. Stroudsburg: Association for Computational Linguistics, 2010. 1–12.

附中文参考文献:

- [42] 邹博伟,钱忠,陈站成,朱巧明,周国栋.面向自然语言文本的否定性与不确定性信息抽取.软件学报,2016,27(2):309–328. <http://www.jos.org.cn/1000-9825/4860.htm> [doi: 10.13328/j.cnki.jos.004860]



钱忠(1989—),男,江苏常熟人,博士,主要研究领域为自然语言处理.



周国栋(1967—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为自然语言处理.



李培峰(1971—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为自然语言处理.



朱巧明(1963—),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为自然语言处理.