多机制整合的中文医疗命名实体识别

王珊珊[1],张焜元[2,3,4],闫蓉[2,3,4*]

[1]延安大学经济与管理学院 陕西, 延安, 716000

[2]内蒙古大学计算机学院 内蒙古, 呼和浩特, 010021

[3]蒙古文智能信息处理技术国家地方联合工程研究中心 内蒙古, 呼和浩特, 010021

[4]内蒙古自治区蒙古文信息处理技术重点实验室 内蒙古, 呼和浩特, 010021

wangshanshan@126.com, 11467705@qq.com, csyanr@imu.edu.cn

摘要

在互联网在线医疗领域，由于大多数患者缺乏医学培训，以及不同学科病理特征的复杂性，医患对话文本中的医学命名实体呈现出长且多词的句法特点，给命名实体识别算法提出了新的挑战。为解决这一问题，本研究融合多个不同粒度的扩张卷积机制，构建了Flat-Lattice-CNN模型。该模型不仅考虑字符和词语的语义信息以及它们的绝对和相对位置信息，还提取跨越不同距离的多个字符/词语的共现依存关系特征，以此提高医学长命名实体的识别精度。实验结果表明，本文提出的模型在所评估数据集的命名实体识别任务上有普遍性的性能提升，尤其是在以长实体为主的中文医疗数据集CTDD上，该模型的$F1$值提升了约2%，具有更优的表现。

关键词： 长且多词医学命名实体 ； 中文医疗命名实体识别 ； 扩张卷积

Infusing multi-schemes for Chinese Medical Named Entity Recognition

Shanshan Wang[1],Kunyuan Zhang[2,3,4],Rong Yan[2,3,4*]

[1]Economics and Management School, Yan'an University, Yanan, 716000, China

[2]College of Computer Science, Inner Mongolia University, Hohhot, 010021, China

[3]National and Local Joint Engineering Research Center of Intelligent Information Processing Technology for Mongolian, Hohhot, 010021, China

[4] Inner Mongolia Key Laboratory of Mongolian Information Processing Technology, Hohhot, 010021, China

wangshanshan@126.com, 11467705@qq.com, csyanr@imu.edu.cn

Abstract

In the field of internet-based healthcare, the complexity of pathology features across various disciplines, coupled with the lack of medical training among most patients, results in medical named entities in doctor-patient dialogue texts exhibiting long and

第二十三届中国计算语言学大会论文集，第382页-第393页，太原，中国，2024年7月25日至28日。
卷1：主会论文

multi-word syntactic patterns, posing new challenges to named entity recognition algorithms. To address this issue, this study integrates multiple dilation convolution mechanisms of different granularities to construct the Flat-Lattice-CNN model. This model not only considers the semantic information of characters and words, as well as their absolute and relative positional information but also extracts relationship features between characters/words spanning different distances to improve the recognition accuracy of long medical named entities. Experimental results show a significant performance boost in the task of recognizing medical named entities on all evaluation dataset, especially on CTDD with a 2% increase in $F1$ score.

Keywords: Long multiword named entities , Chinese Medical Named Entity Recognition , Dilated Convolution

1 引言

命名实体识别(Named Entity Recognition, NER)是自然语言处理的基础任务之一，其主要功能是在文本中识别特定概念的实体，对于机器翻译、信息提取、问答系统等领域至关重要。医学命名实体识别旨在从海量的非结构化医疗数据中提取关键信息，如药物、手术、症状、身体、疾病等，为医学研究发展和智慧医疗系统普及提供基础支持(张帆 and 王敏, 2017)。近年来，随着信息技术的不断进步和人们服务意识的提高，互联网在线医疗正逐渐成为传统面对面医疗的重要补充(罗歆然 et al., 2024)。在互联网在线医疗平台上，诊疗期间的医患交流对话多以文本形式记录并保存。但由于大多数患者缺乏医学培训，以及医学不同学科病理特征的复杂性，他们往往在提供病情和病史时采用较长的表达方式。因此，相关的命名实体表现出长且多词的特征，例如，妇产科领域中“月经过后老是再沥沥拉拉一周左右褐色的分泌物”是一个长且多词的医学症状实体。相比较下，已有研究中的医学命名实体通常以较短和较少词语标记呈现。举例而言，“发烧”被视为一个具有完整且独立医学意义的症状实体，而“断断续续地发烧”也是一个具有完整且独立医学意义的症状实体。尽管它们可能对应着不同的病理情况，然而已有研究却通常将它们等同对待，即“断断续续地发烧”内容中，“断断续续”这一部分词语多数情况下会被忽略标记和识别。

目前，医疗文本命名实体识别的研究主要集中在电子病历、医学文献、医学书籍等领域，而对互联网在线医疗问答文本的研究则较为有限(Li et al., 2020)。虽然已有方法例如BiLSTM-CRF(Luo et al., 2018)、Bert-CRF(Souza et al., 2019) 和LeBert-CRF(Liu et al., 2021)等，在

基金项目：内蒙古自然科学基金(2023MS06023)

医疗命名实体识别领域获得了有效结果，但其应用到互联网在线医疗问答中文文本领域时存在一定局限。具体原因如下：首先，它们主要是针对较短的命名实体识别任务开发的，但在互联网在线医疗医患对话文本中，相关医学命名实体表现出长且多词的句法特点；其次，大多数模型聚焦在英文文本中，而英文和中文的句法特征有较大区别，前者基于字符特征，后者除了字符对词语特征也有较强依赖；最后，虽然命名实体识别普遍被形式化为序列标记问题，但其本质上是对实体内部词之间关系的建模(Li et al., 2022)，然而迄今为止，算法(例如，注意力机制)(Vaswani et al., 2017)更多捕捉的是标记(字符或词语)之间的两两关系信息，而长且多词命名实体中的多词共现依存关系特征仍然缺乏有效的捕获机制。

为解决上述问题，本研究针对互联网医患对话文本，提出了一种整合Flat-Lattice和多粒度扩张卷积网络机制的新模型：Flat-Lattice-CNN。在这个模型中，Flat-Lattice结构简化了模型的复杂性，将其结构设计成了平面形式，不仅降低了模型的复杂度，还充分利用了字符和词级别的信息。虽然Flat-lattice机制可以提取实体中的标记(字符或词语)对之间的长依赖关系，但其主要用于捕获标记之间两两的关系信息。因此，针对长且多词的长医学实体，为了建模实体内多词长距离的共现依存关系特征，本研究整合多粒度扩张卷积机制。实验结果表明，该模型在数据集的命名实体识别任务上有普遍性的性能提升。

2 相关工作

2.1 命名实体识别

命名实体识别（NER）长期以来一直是自然语言处理（NLP）领域的基础任务之一，有着广泛的应用，包括但不限于机器翻译、信息提取、问答系统等。从命名实体识别问题的分析角度，近年来，NER的研究逐渐从最初的平铺式NER(Flat NER)，演变到了嵌套NER(Nested NER)和不连续NER(Discontinuous NER)(Li et al., 2022; Zheng et al., 2019)。

相较于通用领域的命名实体识别，医学命名实体识别任务需要结合医学领域的语言特点和医学语义知识。经过文献梳理，我们发现多数已有医学命名实体工作主要聚焦在输入表示方面，例如Li等采用Lattice-LSTM整合词汇信息，并基于ELMo学习电子病历中的上下文信息(Zhang and Yang, 2018)；(Luo et al., 2018)利用字符串匹配方法将疾病字典与字符进行配对，并提出了结合字典注意力层的BiLSTM-CRF模型；(罗歆然 et al., 2024)提出了一种BERT-BiLSTM-CRF模型，该模型融合了词汇和字根特征，利用BERT模型增强了中文临床记录文本的上下文语义信息。除此之外，也有关注医学文本中的嵌套实体识别的工作(Fei et al., 2021)。近年来，学者们注意到在医学文本中出现不连续命名实体的现象，例如“腿和胳膊疼”。针对这类问题，相关研究提出引入新的标记标签，例如“感觉(疼)”来处理这种不连续实体的情况，这类标签的数量随着语料库的变化，需要及时进行更新和丰富(Dai et al., 2020)。

鉴于先前研究和互联网医患对话文本特点，我们发现长且多词的医学命名实体识别问题尚

未得到重视。例如“断断续续地发烧”中“断断续续”容易被已有医学命名实体识别工作所忽略。考虑到命名实体识别本质上是对实体标记(字符或词)依赖关系的建模，以及中文文本的特点，本研究整合Flat-Lattice和多粒度膨胀卷积机制，提出Flat-Lattice-CNN模型，其中多粒度扩张卷积可有效地捕获跨越短和长距离的多标记(字符或词语)共现依存的关系信息，用以识别长且多词医学命名实体。

2.2 Flat-Lattice

相比英文，中文的命名实体识别更加困难，因为它涉及到词语的分词问题。最近的研究已经证明了栅格结构在利用词语信息和避免分词错误传播方面具有优势。在栅格结构中，可以将一个句子与词典进行匹配，以获取其中的潜在词语，这些词语对于命名实体识别可能非常重要。具体而言，栅格中的每个节点都是标记(一个字符或一个潜在词语)，其中，词语并非按顺序排列，而是由词语的第一个字符和最后一个字符决定其位置。基于栅格的模型使用注意力机制在每个位置融合可变数量的词语节点。然而，由于这种栅格结构的动态性，相关方法不能充分利用GPU的并行计算。因此，李等人提出了Flat-Lattice，该模型基于Transformer将栅格结构转换为平面格(flat-lattice)，通过设计的头(head)和尾(tail)标记的位置编码，该模型可无损复原栅格结构。因此，可以直接使用Transformer来完全模拟栅格输入。不仅如此，Transformer的自注意力和全连接机制使字符可以直接与任何潜在词进行交互，并模拟序列中标记之间的不同距离依赖关系(Li et al., 2020)。

2.3 CNN

卷积神经网络(Convolutional Neural Networks,CNN)(Zeng et al., 2014)是一种结构特殊的人工神经网络，主要应用在视频分类、图像识别、图像标注、人脸检测、信息抽取、文本处理和语音识别等方面，在实际应用中达到很好的效果。卷积神经网络一般由输入层、卷积层、池化层、激活层以及顶端的全连接层、损失函数层组成。卷积层的目的是提取不同特征，位置较低的卷积层提取低级特征，位置较高的卷积层提取高级特征。其中，卷积层的卷积核感受视野范围由膨胀率控制，即膨胀率越高，卷积核的感受视野越大，如果是在文本处理领域，则能建模跨越长距离的多词共现依存关系特征。

3 Flat-Lattice-CNN模型

本文提出的Flat-Lattice-CNN模型主要由三部分组成，分别是输入层、Self-Attention+FFN层和标签预测层，整体结构如图1所示。首先，利用Flat-Lattice Transformer获取输入句子的标记(字符或词语)嵌入向量和标记的绝对、相对位置向量，进一步在此基础上采用不同膨胀率的卷积核将跨越不同距离的多词共现依存关系的特征转化成向量，然后融合这三类型向量，得到最终的输入表示。其次,将输入层得到的向量输入至Self-Attention+FFN层，层

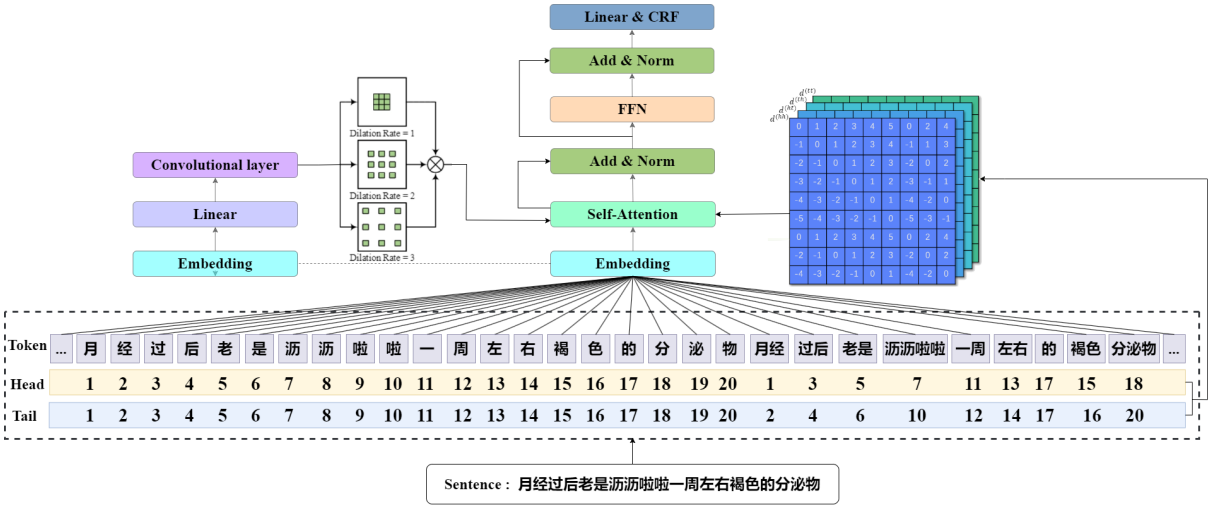

图 1: FLAT-Lattice-CNN模型的整体架构图

中使用Self-Attention和FFN学习标记(字符或词语)的上下文环境，然后使用残差结构得到最终的语义特征。最后,将最终的语义特征输入到以CRF为函数的标签预测层中。接下来将详细说明各个部分。

3.1 Flat-Lattice Transformer

Flat-Lattice Transfomer先将字符与词汇表生成的格子展开为扁平化span序列，之后采用Transformer融合位置信息得出每个标记的嵌入向量。具体的，每一个span对应一个标记(字符或词语)、一个头(head)和一个尾(tail)的位置信息，就像图1虚线框中所示。标记是一个字符或单词。头和尾指示了标记在原序列中第一个和最后一个字符的位置索引，也就是它们表示了标记在格子中的位置。Transformer采用全连接的自注意力机制和前馈网络来建模序列中的标记(字符或词语)之间的依赖关系。

为了保留位置信息,Flat-Lattice Transformer为序列中的每个标记引入了位置表示，形成对应的每个span。具体而言，为每一个标记（字符或单词）分配了两个位置索引：头位置和尾位置，通过这些位置索引，可以从一个标记序列中重构出一个格子。输入序列中的任意两个SPANx_i和x_j，它们之间存在相交、包含和分离三种关系，由span的头和尾共同决定。因此，这种位置机制可以同时表示两个标记绝对和相对位置信息，任意两个SPANx_i和x_j的相对位置可

以用公式1计算：

$$\begin{aligned} d_{ij}^{(hh)} &= head\left[i\right] - head\left[j\right] \\ d_{ij}^{(ht)} &= head\left[i\right] - tail\left[j\right] \\ d_{ij}^{(th)} &= tail\left[i\right] - head\left[j\right] \\ d_{ij}^{(tt)} &= tail\left[i\right] - tail\left[j\right] \end{aligned} \tag{1}$$

其中，d表示头部之间的距离i和j具有类似的含义。

SPAN的相对位置编码是四个距离的简单非线性变换，计算方法如2所示：

$$R_{ij} = \mathrm{ReLU}\left(W_r\left(P_{d_{ij}^{(hh)}} \oplus P_{d_{ij}^{(th)}} \oplus P_{d_{ij}^{(ht)}} \oplus P_{d_{ij}^{(tt)}}\right)\right) \tag{2}$$

其中Wr是可学习参数，$\bigoplus$表示串联运算符，pd的计算方法如Vaswani等人所述(Vaswani et al., 2017)，可通过公式3来计算：

$$\begin{aligned} P_d^{(2k)} &= \sin\left(d/10000^{2k/d_{model}}\right) \\ P_d^{(2k+1)} &= \cos\left(d/10000^{2k/d_{model}}\right) \end{aligned} \tag{3}$$

其中，d和k表示位置编码的维度索引。然后，使用自注意力的变体来利用相对SPAN的位置编码，以进一步提高模型对序列信息的建模能力。通过融合的标记(字符或词语)和它们的相对、绝对位置信息，并经过transformer机制后，得到了输入序列中每个span的嵌入向量w_F。

3.2 Dilated Convolution CNN

在前述嵌入向量w_F基础上，本研究引入了不同膨胀率的卷积机制，以此获取实体内跨越不同距离的多词共现依存关系特征向量w_c，然后与w_f拼接成为最终的输入嵌入向量$[w_f; w_c]$。如公式4所示，卷积机制具有局部感受视野和平移不变性的机制，局部感受野是指机制能够专注于局部特征的提取，而不需要考虑整个输入的复杂性；平移不变性指的是是指特征的识别不受其绝对位置的影响。 Dilated Convolution CNN结构包括以下几个关键组件：

1.基础卷积层：首先通过一个具有可选丢弃率的一维卷积层来对输入进行特征提取。这有助于将输入数据映射到一个更高维度的表示空间，并通过RELU激活函数进行非线性转换。这一步旨在为后续的多粒度扩张卷积层提供一个丰富的输入表示。

2.多粒度扩张卷积(Li et al., 2022)：为了捕获多词的共现依存关系信息，本文引入了多个具有不同扩张率的卷积核。每个卷积核都具有自己的感受野大小，使其能够有效地捕获不同范围内的语义信息。具体地，假设以一个变量a来衡量空洞卷积的扩张系数，则加入空洞之后的实际卷积核尺寸与原始卷积核尺寸之间的关系为公式4：

$$K = k + \left(k - 1\right)\left(a - 1\right) \tag{4}$$

其中，k为原始卷积核大小，a为卷积扩张率(dilation rate)；K为经过扩展后实际卷积核大小。卷积后得到的尺寸用公式5计算：

$$W = \frac{w-k+2p}{s}+1 = \frac{w-k-(k-1)(a-1)+2p}{s}+1 \tag{5}$$

其中，w为输入尺寸，k为原始卷积核大小，a为卷积扩张率，s 为stride，p是padding。

3.残差连接(He et al., 2016)：为了确保信息的流动和保留原始输入的重要信息，本文在多粒度扩张卷积之后采用了残差连接的策略。

3.3 Self-Attention+FFN

本研究在整合Flat-Lattice Transformer和多通胀率的卷积机制基础上获取了输入序列的嵌入表示，之后将嵌入表示的序列输入到自注意力(Self-Attention)和全连接层(FFN)，使得模型在输入层基础上进一步提升处理每个元素考虑整个序列上下文信息的能力，从而更好地理解元素与序列中其他元素之间的关系。

3.4 预测层

标签预测层采用的是条件随机场模型 (Condi-tional Random Field,CRF)(Lafferty et al., 2001)。CRF是一种判别式的无向图模型，通过研究标签之间的关系，获得全局最优的标签序列。

4 实验结果与分析

4.1 实验数据

为验证本文模型在中文命名实体识别任务中的有效性，在IMCS-2021(Chen et al., 2023)和CTDD(Wang et al., 2023)两个公共中文医疗NER数据集进行实验，两个数据集的描述信息如表 1所示。CTDD是妇产科领域互联网医患对话中文数据集，IMCS-NER则是儿科领域的在线医患对话中文数据集。如表1所示，CTDD的实体平均长度为4.33，IMCS-2021则是2.62，因此可以得出CTDD的长实体比例比IMCS-2021更高。

IMCS-2021数据集涵盖症状、检查、药物名称等5种实体类型，本文从中随机抽取73,603条数据作为训练集，12,517条数据作为验证集，12,332条数据作为测试集。CTDD数据集包含时间、疾病和症状等6种实体类型，本文随机抽取21,633条作为训练集，3,757条作为验证集，4,847作为测试集。其中，CTDD数据集实体类型占比如图 2所示。

在本文命名实体识别实验评估任务中，参数设置如表 2所示。

	CTDD	IMCS-NER
所有命名实体的数量	63560	74698
实体平均长度	4.33	2.62
总字符数	1,700,392	1,621,161
标记字符与总字符的比例	16.2%	12.1%
对话的平均字数	713.55	589.04

表 1: CTDD数据集与IMCS-NER数据集的描述信息

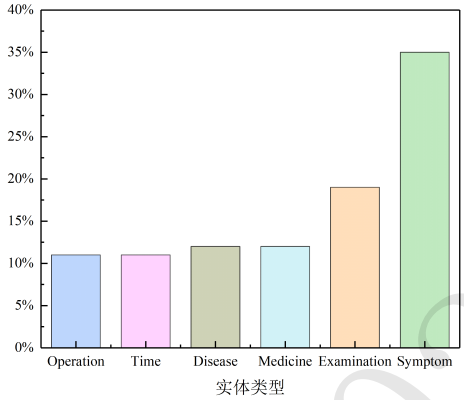

图 2: CTDD数据集实体类型分布图

4.2 评价指标

本文实验结果使用$F1$值作为评价指标，计算如公式6所示:

$$\begin{aligned} P &= \frac{\text{预测正确的实体集}}{\text{得到的实体总数}} \times 100\% \\ R &= \frac{\text{预测正确的实体集}}{\text{数据集实体数}} \times 100\% \\ F_1 &= \frac{2 \times P \times R}{P + R} \times 100\% \end{aligned} \tag{6}$$

4.3 基线模型

(1)BiLSTM-CRF(Luo et al., 2018)模型采用双向长短期记忆网络(Bidirectional Long Short-Term Memory，BiLSTM)来捕获输入序列中的上下文信息，并结合条件随机场(Conditional Random Field，CRF)来对序列标注问题进行建模。

(2)Bert-CRF(Souza et al., 2019)模型结合了预训练的 BERT 模型和条件随机场(constant ratefactor，CRF)，利用 BERT 模型强大的上下文理解能力，对输入序列进行表示学习，并通过 CRF 层进行标签序列的全局优化。

参数	数量
学习率	6e-4
batch	10
epoch	50
注意力头的个数	8
注意力头的维度	20
输入维度	160
通道数	160
Dropout	0.5

表 2: 参数设置

(3)LeBert-CRF(Liu et al., 2021)模型在 Bert-CRF 模型的基础上进行了改进，引入了领域专属的预训练方法。

(4)Lattice LSTM(Zhang and Yang, 2018)模型是一种基于格子结构的神经网络模型，它在模型的输入端引入了字符级别的信息，并通过格子结构的设计，使得模型能够更好地捕获单词内部字符之间的依赖关系。

4.4 实验结果分析

为更好观察本文模型对中文医疗命名实体识别的效果，分别用BiLSTM-CRF、Lattice LSTM、Bert-CRF、LeBert-CRF、FLAT-Lattice-Transformer模型作为对照组，分别在IMCS-2021和CTDD两个公共中文医疗NER数据集进行实验。对照组模型参数均为默认值，不同模型的中文命名实体识别效果如表3所示。

	CTDD			IMCS-2021			weibo		
模型	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)
Bert-CRF	53.09	52.75	52.92	88.46	92.35	90.37	-	-	-
LeBert-CRF	52.63	54.25	53.42	86.53	92.91	89.60	-	-	-
BiLSTM-CRF	46.26	50.34	48.21	85.67	88.72	87.61	-	-	-
Lattice LSTM	**57.14**	55.17	56.14	89.37	90.84	90.10	53.04	62.25	58.79
FLAT-Lattice Transformer	53.62	56.47	55.14	88.83	93.43	91.07	69.84	77.37	73.41
Flat-Lattice-CNN (ours)	55.63	**59.10**	**57.31**	**89.44**	**93.51**	**91.43**	**70.29**	**79.69**	**74.69**

表 3: 在CTDD、IMCS、weibo数据集上的实验结果

虽然长且多词的实体是医学领域的文本特点，但是其它领域也可能有长实体的存在，因此本文对微博中文数据集在算法FLAT-Lattice Transformer和Flat-Lattice-CNN等进行了对比实验评估，如表3。由表3可以观察到，该模型有普遍性的性能提升，尤其时在长实体数据集上，

$F1$值提升了2%，这也在一定程度上证明了本文模型的有效性。

为更好观察本文中多粒度扩张卷积的对中文医疗命名实体识别效果的影响，λ为膨胀率，即Dilation分数，本文分别建立了λ=[1]，λ=[1,2]，λ=[1,2,4]作为对照组模型，其他参数均为默认值，不同$dilation$($epoch$=15情况下的在CTDD数据集的命名实体识别效果如表4所示，由表中可以观察到膨胀率λ=[1,2,4]时$F1$分数和准确率(ACC)值均有所提高。在模型中增加膨胀率4的卷积核，我们发现实验性能比低膨胀率的时候性能提升。

Dilation(λ)	$F1$(%)	ACC(%)
[1]	53.40	86.53
[1,2]	54.50	87.61
[1,2,4]	55.35	87.75
[1,2,4,8]	52.77	87.70

表 4: 膨胀率的对比

句子	超过38.5度才给孩子用退烧药
Bert-CRF	超过38.5度(Symptom)才给孩子用退烧药(Drug)
Lattice LSTM	超过38.5度(Symptom)才给孩子用退烧药(Drug)
Ours	超过38.5度(Symptom)才给孩子用退烧药(Drug)

表 5: IMCS-2021数据集命名实体识别效果实例

进一步地，本文列举出一个在以短实体为主的IMCS-2021数据集上的实例，如表5所示，其中Drug代表药物，退烧药是一个药物命名实体，Symptom代表症状。由表5可以观察到，Lattice LSTM模型将“38.5度”识别为短的单独的症状命名实体，而本文模型将“超过38.5度”成功识别为一个较长的症状命名实体，由表5可以直观地观察到本文模型对长实体识别改进的有效性。

由于本文采用了卷积层来捕获实体内多标记共现依存关系特征，在实验中我们观察到，模型捕获信息的速度更快，在$epoch$=15的情况下在CTDD数据集中$F1$分数的训练结果已经达到了52%（约为最终结果的91.2%）而FLAT-Lattice-Transformer的$F1$分数达到了0.48%（约为最终结果的87.2%），模型推理速度对照结果如图3所示。

5 结束语

本文介绍了一种基于 Flat-Lattice-CNN 的中文医疗命名实体识别方法，通过结合 Flat-Lattice 结构和不同粒度的膨胀卷积机制，有效地捕获了医疗文本中长且多词的命名实体中的多词共现依存关系信息。一定程度上克服了以注意力机制为基础的算法在这方面的局限。实验

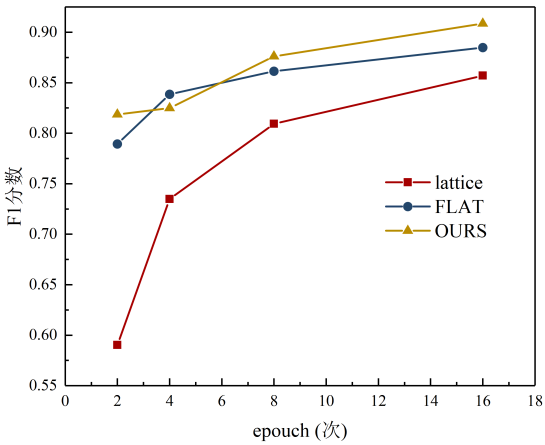

图 3: 模型推理速度对照结果图

结果表明，相比传统方法，本文模型在$F1$值和收敛速度上均有一定的提升，丰富了医疗信息处理领域的思路和方法。将来，我们会探索长且多词命名实体识别效果的语言学理论理解和效果的进一步优化，以应对更加复杂和多样化的医疗文本数据。

致谢

本项研究获得了国家自然科学基金(编号:72261032)、内蒙古自然科学基金(编号:2023MS06023)和陕西社会科学基金(编号:2022N005)资助。

参考文献

Wei Chen, Zhiwei Li, Hongyi Fang, Qianyuan Yao, Cheng Zhong, Jianye Hao, Qi Zhang, Xuanjing Huang, Jiajie Peng, and Zhongyu Wei. 2023. A benchmark for automatic medical consultation system: frameworks, tasks and datasets. *Bioinformatics*, 39(1):btac817.

Xiang Dai, Sarvnaz Karimi, Ben Hachey, and Cecile Paris. 2020. An effective transition-based model for discontinuous ner. *arXiv preprint arXiv:2004.13454*.

Hao Fei, Donghong Ji, Bobo Li, Yijiang Liu, Yafeng Ren, and Fei Li. 2021. Rethinking boundaries: End-to-end recognition of discontinuous mentions with pointer networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 12785–12793.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

John Lafferty, Andrew McCallum, Fernando Pereira, et al. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Icml*, volume 1, page 3. Williamstown, MA.

Xiaonan Li, Hang Yan, Xipeng Qiu, and Xuanjing Huang. 2020. Flat: Chinese ner using flat-lattice transformer. *arXiv preprint arXiv:2004.11795*.

Jingye Li, Hao Fei, Jiang Liu, Shengqiong Wu, Meishan Zhang, Chong Teng, Donghong Ji, and Fei Li. 2022. Unified named entity recognition as word-word relation classification. In *proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 10965–10973.

Wei Liu, Xiyan Fu, Yue Zhang, and Wenming Xiao. 2021. Lexicon enhanced chinese sequence labeling using bert adapter. *arXiv preprint arXiv:2105.07148*.

Ling Luo, Zhihao Yang, Pei Yang, Yin Zhang, Lei Wang, Hongfei Lin, and Jian Wang. 2018. An attention-based bilstm-crf approach to document-level chemical named entity recognition. *Bioinformatics*, 34(8):1381–1388.

Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2019. Portuguese named entity recognition using bert-crf. *arXiv preprint arXiv:1909.10649*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Shanshan Wang, Yajing Yan, Rong Yan, Ting Li, Kaijie Ma, and Yani Yan. 2023. A chinese telemedicine-dialogue dataset annotated for named entities. *BMC Medical Informatics and Decision Making*, 23(1):264.

Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. 2014. Relation classification via convolutional deep neural network. In *Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers*, pages 2335–2344.

Yue Zhang and Jie Yang. 2018. Chinese ner using lattice lstm. *arXiv preprint arXiv:1805.02023*.

Changmeng Zheng, Yi Cai, Jingyun Xu, HF Leung, and Guandong Xu. 2019. A boundary-aware neural model for nested named entity recognition. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics.

张帆 and 王敏. 2017. 基于深度学习的医疗命名实体识别. 计算技术与自动化, 36(1):123–127.

罗歆然, 李天瑞, and 贾真. 2024. 基于自注意力机制与词汇增强的中文医学命名实体识别. 计算机应用, 44:385–392.