

doi:10.6041/j.issn.1000-1298.2022.03.031

兽药致病命名实体 Att - Aux - BERT - BiLSTM - CRF 识别

杨璐¹ 张恬¹ 郑丽敏^{1,2} 田立军¹

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 食品质量与安全北京实验室, 北京 100083)

摘要: 针对兽药致病知识图谱构建过程中,关于兽药命名实体识别使用传统方法依赖人工设计特征耗时耗力以及兽药致病语料数据量较少的问题,提出一种引入注意力机制(Attention)与辅助层分类(Auxiliary layer)相结合兽药文本命名实体识别模型(Att - Aux - BERT - BiLSTM - CRF)。通过 BERT 预处理模型进行文本向量化,然后连接双向长短期记忆网络(Bi-directional long-short term memory, BiLSTM);引入注意力机制,将模型的 BERT 层输出用作辅助分类层, BiLSTM 层输出作为主分类层(Mainlayer),通过注意力机制组合以提高整体性能;最后输入条件随机场(Conditional random field, CRF),构建端到端的适合于兽药领域实体识别的深度学习模型框架。实验选取兽药文本共 10 643 个句子、485 711 个字符,针对动物、药物、不良反应、摄入方式 4 类实体进行识别。实验结果表明,本文模型能有效地辨别兽药致病文本中的实体,识别的 F1 值为 96.7%。

关键词: 兽药致病; 命名实体识别; 注意力机制; BERT; 深度学习

中图分类号: TP391.1 文献标识码: A 文章编号: 1000-1298(2022)03-0294-07

OSID:



Recognition of Animal Drug Pathogenicity Named Entity Based on Att - Aux - BERT - BiLSTM - CRF

YANG Lu¹ ZHANG Tian¹ ZHENG Limin^{1,2} TIAN Lijun¹

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Beijing Laboratory of Food Quality and Safety, Beijing 100083, China)

Abstract: In order to solve the problems that traditional methods of veterinary drug named entity recognition rely on artificial design features, which is time-consuming and labor-consuming, and the amount of veterinary drug pathogenic corpus data is less in the process of building veterinary drug pathogenic knowledge graph, a method based on Att - Aux - BERT - BiLSTM - CRF of veterinary drug text named entity recognition model was proposed, which combined BERT - BiLSTM - CRF models by introducing attention mechanism and auxiliary classification layer. The text was vectorized by the BERT preprocessing model, and then connected to bi-directional long-short term memory network. The auxiliary classification mechanism was introduced, the output of the BERT layer was used as the auxiliary classification layer, and the output of the BiLSTM layer was used as the main classification layer. The attention mechanism was proposed to combine auxiliary classification layer with main classification layer to improve the overall performance. Finally, it was sent to conditional random field to construct an end-to-end deep learning model framework suitable for veterinary drug name entity recognition. In the experiment, totally 10 643 sentences and 485 711 characters of veterinary drug text were selected to identify four kinds of entities: drug, adverse effect, intake mode, aimal. The results showed that the model can effectively identify the entities in the veterinary drug pathogenic text, and the F1 value of recognition was 96.7%.

Key words: veterinary drug pathogenicity; named entity recognition; attention mechanism; BERT; deep learning

收稿日期: 2021-02-02 修回日期: 2021-04-15

基金项目: 北京市现代农业产业技术体系创新团队项目(BAIC02-2020)和国家重点研发计划项目(2017YFC1601803)

作者简介: 杨璐(1967—),女,副教授,博士,主要从事计算机网络工程与人工智能研究,E-mail: yanglumail@263.net

0 引言

兽药残留会对人体健康造成巨大危害^[1]。文献[2]提出了一种根据兽药通过环境间接接触人类的可能性及其毒性概况来对兽药进行排序的方法,但是由于许多数据缺失,所以准确性不高。如何快速准确地从兽药致病文本中获得相关信息,是构建兽药致病知识图谱中的关键环节。兽药致病命名实体识别是兽药致病信息抽取的重要任务之一,目的是从兽药致病非结构化文本数据中获得兽药名称、实验动物、不良反应等信息。

命名实体识别的方法包含基于规则、机器学习和深度学习三大类。基于规则的命名实体识别的方法耗时耗力且可移植性差^[3]。基于机器学习的方法将命名实体识别问题转化为序列标注问题,如文献[4]采用条件随机场(Conditional random field, CRF)进行命名实体识别,文献[5]采用 CRF 对农业病虫害实体进行识别。但是传统的机器学习方法仍然需要人工选取特征^[6]。

基于深度学习的命名实体识别方法广泛应用于渔业^[7]、农业病虫害^[8]、电子病历^[9]、医学文献^[10]和新闻社论^[11]等领域。文献[12]采用 LSTM 和 CRF 对文献中药物名称进行命名实体识别,在 DDI2011 语料库上精度达到 93.26%。文献[13]首次联合采用 BiLSTM 和 CRF 进行命名实体识别,获得了 F1 值为 90.10%。基于 BiLSTM 方法学习文本的长距离依赖,相较于 LSTM 方法,可以充分结合上下文信息,同时 CRF 可以获得相邻标签之间的依赖关系。双向编码表征模型(Bidirectional encoder representations from transformers, BERT)^[14]的出现提升了命名实体识别的效率,文献[15]采用双向编码表征-双向门限循环-条件随机场模型(Bidirectional encoder representations from transformers - Bidirectional gated recurrent unit - Conditional random field, BERT - BiGRU - CRF)进行命名实体识别,解决了传统的词向量表示方法只是将字映射为单一向量,无法表征字的多义性的问题,在 MSRA 语料上的 F1 值达到 95.43%。文献[16]采用 BERT - BiLSTM - CRF 模型对人民日报语料库进行命名实体识别, F1 值为 95.67%。基于深度学习的方法命名实体识别任务转化为字级别的序列标注任务,减少了特征工程的工作量,同时结合了上下文信息,提高了模型的性能,但是需要大量的标注语料。

基于注意力机制的神经网络是深度学习研究的一个热点。受人类观察图像时会对特定部分集中注意力的启发,注意力机制首先应用在图像处理领域。

文献[17]将模型最后层的输出和中间层的输出结合起来,在图像分类领域和图像分割领域^[18]取得了较好的结果。注意力机制^[19]在自然语言处理领域也得到广泛应用。文献[20]提出使用注意力机制将字级别和词级别的嵌入结合起来,可以获得更好的结果。文献[21]分析使用不同的注意力计算函数对于命名实体识别模型的影响。

发布于有害物质数据库的兽药致病文本表达不规范,没有特定的规则,且部分实体词长度较长、使用缩写简称等情况,导致实体识别难度较大。兽药致病实体词汇较长是多个一般词汇组合导致的,同一实体的缩写简称也表现为多种形式。这就需要结合上下文,充分利用词汇的句法和语法信息,而不能仅仅依靠基于统计的方法来设计规则。针对这一问题,本文采用预训练语言模型的方法,学习包含兽药致病领域上下文语义信息的词向量,提出基于注意力机制的端到端的 Att - Aux - BERT - BiLSTM - CRF 神经网络模型。首先将兽药致病领域命名实体识别任务转换为字级别的序列标注任务,然后基于 BERT 进行字向量的提取,使用 BERT 中文预训练兽药致病领域词向量作为 BiLSTM 网络的输入,将模型的 BERT 层输出用作辅助分类层, BiLSTM 层输出作为主分类层,引入皮尔逊相关系数作为注意力值计算函数,计算 2 个层的相似度。最后输入条件随机场,通过相邻标签的依赖关系确定输出序列。

1 数据获取与标注

1.1 数据获取

本文通过爬虫获得兽药致病数据集。研究兽药残留致病信息的数据主要来自于两个网站。有害物质数据库(HSDB)(<https://www.nlm.nih.gov/toxnet/index.html>)是一个毒理学数据库,着眼于分析存在潜在危险的化学物质。它提供了兽药对于人体健康的影响和非人类毒性摘要等信息。粮农组织/世卫组织食品添加剂联合专家委员会(JECFA)网站(<https://apps.who.int/food-additives-contaminants-jecfa-database/search.aspx>)可搜索的数据包含 JECFA 对风味、食品添加剂、污染物、有毒物质和兽药进行的所有评估的摘要。每个摘要均包含基本化学信息、最新报告、专著链接、规格数据库以及 JECFA 评估的历史记录。可通过部分名称或编号来搜索数据。

《食品安全国家标准 食品中兽药最大限量》中规定的兽药共包含 3 类:可以使用但是有残留限制的兽药、不需要设定残留限制的兽药、不得检出的兽药。从兽药残留致病角度考虑,选择第 1 类和第

3 类兽药共计 113 种。利用爬虫在这两个网站中以兽药名称为关键词进行检索,获得相关的兽药致病数据,共计 10 643 个句子,485 711 个字符。

1.2 数据标注

语料库共包含实体 14 189 个,具体分布如表 1 所示。使用 BIO 标注方法进行标注,B 表示实体的开始,I 表示实体的内部,O 表示非实体。

表 1 标注数据数量

Tab.1 Label data quantity

标注内容	标注数量/条
实体	14 189
药物	4 067
不良反应	4 281
动物	4 038
摄入方式	1 803

2 模型框架

本文提出的 Att - Aux - BERT - BiLSTM - CRF 模型由 BERT 模块、BiLSTM 模块、基于注意力机制的辅助分类器和 CRF 模块共 4 个模块组成,模型结构如图 1 所示。首先通过 BERT 预训练模型获取字向量,提取文本中的重要特征;然后使用 BERT 中文预训练兽药致病领域词向量作为 BiLSTM 网络的输入,通过 BiLSTM 层学习上下文特征;将模型的 BERT 层输出用作辅助分类层,BiLSTM 层输出作为主分类层,接着将这两部分的结果通过注意力机制加权融合在一起,作为新的特征,输入到 CRF 层中,得到最后的输出序列。

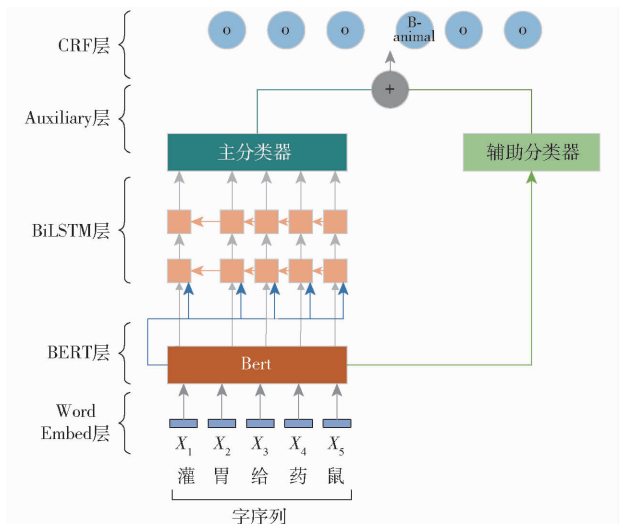


图 1 Att - Aux - BERT - BiLSTM - CRF 模型结构

Fig.1 Att - Aux - BERT - BiLSTM - CRF model structure

2.1 BERT 层

在自然语言预处理领域存在 Word2Vec^[22]、ELMo^[23]、GPT^[24] 等多种预训练模型。BERT 融合了

这些模型的优点并克服了他们的缺点。BERT 使用遮蔽语言模型 (Masked language model, MLM) 和下一句预测来进行预训练。这样得到的向量不仅包含隐含的上下文信息,还包含句子级别的特征。图 2 是 BERT 结构示意图。图中 E_N 、 T_N 、Trm 分别代表输入的嵌入层、输出的 BIO 标志、Transformers 编码器。

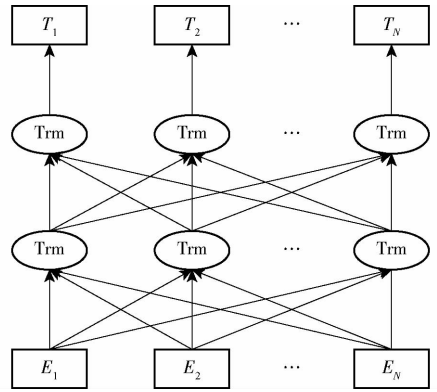


图 2 BERT 结构示意图

Fig.2 BERT structure diagram

BERT 模型中 Transformer 采用自注意力机制和全连接层处理输入的文本。Transformer 采用的多头注意力机制,可以获得多个维度的信息。本文模型将 BERT 的输出结果输入到 BiLSTM 层中。

2.2 BiLSTM 层

BiLSTM 层获得 BERT 输出的向量,从而获得更加全面的语义信息。LSTM 可以有效解决循环神经网络 (Recurrent neural network, RNN) 中的梯度消失或梯度爆炸的现象^[25]。本文采用了双向 LSTM (BiLSTM) 结构。例如,兽药实体“马拉硫磷”,LSTM 仅可以获得“拉”的前一个字“马”的特征信息,而 BiLSTM 还可以获得“拉”的后一个字“硫”的特征信息,从而更好地捕获上下文信息。即对于每一个句子,分别计算从左至右和从右至左两个顺序的表示 \vec{h}_1 和 \overleftarrow{h}_2 ,然后将其拼接得到 t 时刻 BiLSTM 的输出 $[\vec{h}_1; \overleftarrow{h}_2]$ 。LSTM 的公式为

$$F_t = \sigma(W_f[H_{t-1}, X_t + b_f]) \tag{1}$$

$$I_t = \sigma(W_i[H_{t-1}, X_t + b_i]) \tag{2}$$

$$\tilde{C}_t = \tanh(W_c[H_{t-1}, X_t + b_c]) \tag{3}$$

$$C_t = F_t C_{t-1} + I_t \tilde{C}_t \tag{4}$$

$$O_t = \sigma(W_o[H_{t-1}, X_t + b_o]) \tag{5}$$

$$H_t = O_t \tanh(C_t) \tag{6}$$

式中 σ ——sigmoid 激活函数

\tanh ——双曲正切激活函数

F_t, I_t, C_t, O_t ——在 t 时刻遗忘门、输入门、记忆细胞和输出门

W_f, W_i, W_c, W_o ——不同门对应的权重矩阵

b_f, b_i, b_c, b_o ——偏置向量

\tilde{C}_t ——输入的中间状态向量

X_t —— t 时刻的输入向量

H_t —— t 时刻的输出向量

BERT 层输出向量通过 BiLSTM 层输出 $[\vec{h}_1; \overleftarrow{h}_2]$, 并将其作为 Attention 层主分类器的输入。

2.3 Attention 层

在计算机视觉领域, 已知将模型的中间层输出结果和最后输出结果相结合可以改善图像分割的性能。这是因为模型的中间层输出结果和最后输出结果对于图像分割都有作用, 所以将中间层的输出结合起来进行图像分割。本文将 BERT 层输出结果作为辅助分类器, BiLSTM 层输出结果作为主分类器。BERT 可以获得丰富的语义信息, BiLSTM 可以获得长距离的上下文信息。将两者结合起来, 从而更好的进行序列标注。

注意力机制起源于对人类注意力的模仿, 其本质是通过注意力分配参数筛选出特定的信息, Attention 层的主要作用是衡量特征权重。在之前的 Aux - BERT - BiLSTM - CRF 模型中, BERT 层输出结果(辅助分类器)和 BiLSTM 层输出结果(主分类器)具有相同的权重, 这样会产生两个弊端: BERT 层输出的内容会被稀释掉; 语义向量无法表示整个序列的所有信息。使用 Attention 机制可以计算两者的相关性, 动态分配权重, 让更重要的特征占据更大的权重, 从而获得对于序列标注更为重要的信息。本文将 BERT 层输出结果(辅助分类器)和 BiLSTM 输出结果(主分类器)通过 Attention 机制结合起来。

在注意力层中, 引入注意力机制来计算辅助分类器和主分类器的相似性得分。注意力值计算函数(Score 函数)是衡量 BERT 层输出向量和 BiLSTM 输出向量之间相关性的得分函数。文献[22]设计了余弦距离、欧氏距离和感知机 3 种方案作为得分函数。皮尔逊相关系数、马氏距离分别是余弦距离和欧氏距离的改进, 本文设计了皮尔逊相关系数、马氏距离、感知机 3 种方案, 其 Score 函数计算公式为

$$\text{Score}(\mathbf{h}_t, \mathbf{h}_s) =$$

$$\begin{cases} \mathbf{W} \cdot \frac{(\mathbf{h}_t - \bar{\mathbf{h}}_t) \cdot (\mathbf{h}_s - \bar{\mathbf{h}}_s)}{\sqrt{(\mathbf{h}_t - \bar{\mathbf{h}}_t)^2} \sqrt{(\mathbf{h}_s - \bar{\mathbf{h}}_s)^2}} & (\text{皮尔逊相关系数}) \\ \mathbf{W} \cdot (\mathbf{h}_t - \mathbf{h}_s)^T \cdot \mathbf{S}^{-1} \cdot (\mathbf{h}_t - \mathbf{h}_s) & (\text{马氏距离}) \\ \tanh(\mathbf{W}[\mathbf{h}_t; \mathbf{h}_s]) & (\text{感知机}) \end{cases} \quad (7)$$

式中 \mathbf{h}_t ——BERT 层输出的结果, 即辅助分类器

\mathbf{h}_s ——BiLSTM 层输出的结果, 即主分类器

\mathbf{W} ——权重矩阵

$\bar{\mathbf{h}}_s, \bar{\mathbf{h}}_t$ ——分类器平均值

\mathbf{S} ——协方差矩阵

皮尔逊相关系数是数据中心化后的两个多维向量的夹角余弦。两个向量越相似, 皮尔逊相关系数和感知机的计算结果得分越大。马氏距离是度量学习中常用的距离指标, 用于评定数据之间的相似度。使用 Score 函数得到两个层的特征权重。然后通过这两个粒度的向量特征相乘得到新的特征输入到 CRF 层中。

2.4 CRF 层

在命名实体识别任务中, BiLSTM 可以处理长距离的文本信息, 但是无法处理相邻标签之间的依赖关系。条件随机场(Conditional random field, CRF)可以通过临近标签获得一个最优的预测序列。

对于输入序列 $X = (X_1, X_2, \dots, X_n)$, 提取特征得到输出矩阵 $\mathbf{P} = [P_1 \ P_2 \ \dots \ P_n]$ 。对于预测序列 $Y = (y_1, y_2, \dots, y_n)$, 其分数函数公式为

$$S(X, Y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=1}^n P_{i, y_i} \quad (8)$$

式中 $A_{y_i, y_{i+1}}$ ——标签 y_i 转移到标签 y_{i+1} 的分数

P_{i, y_i} ——字符被预测为第 y_i 个标签的分数

最后使用 softmax 层计算出所有可能标签的概率, 输出一个得分最高的标记序列。

3 实验

实验采用 Pytorch 1.7.1 框架, 运行环境为 RTX 2080Ti GPU, 内存为 11 GB。

3.1 实验设置

参数配置如表 2 所示, 为解决过拟合问题, 引入 dropout^[26] 机制, dropout 值为 0.5。优化算法为 Adam^[27]。

表 2 参数配置

Tab. 2 Parameter configuration

参数	数值
Transformer 层	12
字向量维度	100
隐藏单元数	128
Dropout	0.5
学习率	0.000 01
注意力输出维度	100
批量大小	32
迭代周期	15

3.2 评价方法

采用准确率、召回率、F1 值作为评估指标^[28]。

3.3 实验结果

在标注数据集上验证模型的性能。语料库中的

训练集、测试集按照 8:2 进行分配,数据集之间无重叠,因此实验结果可以作为实体识别效果的评价指标。训练集和测试集实体数量见表 3。

表 3 数据集实体信息

Tab. 3 Dataset entity information

类别	名称	训练集实体 数量/条	测试集实体 数量/条
药物	Medicine	3 246	821
不良反应	Adverse	3 400	881
动物	Animal	3 288	750
摄入方式	Get	1 446	357

3.3.1 不同模型的性能比较

在不同的模型上进行实验,模型包括 BiLSTM-CRF^[13]、BERT-BiLSTM-CRF^[15] 以及本文提出的 Aux-BERT-BiLSTM-CRF 和 Att-Aux-BERT-BiLSTM-CRF。Aux-BERT-BiLSTM-CRF 模型将 BERT 层输出的结果作为辅助分类器,BiLSTM 层输出的结果作为主分类器,将两者拼接起来。Att-Aux-BERT-BiLSTM-CRF 模型与前者不同之处在于,使用注意力机制,选择皮尔逊相关系数作为注意力值计算函数公式,将两者动态结合起来。

通过表 4 可知,BiLSTM-CRF 模型 F1 值为 94.1%。BERT-BiLSTM-CRF 模型相较于 BiLSTM-CRF 模型,由于引入 BERT 可以获得丰富的上下文信息,模型 F1 值提高了 1.1 个百分点。将 BERT 输出的结果和 BiLSTM 输出的结果通过注意力机制结合起来,通过学习判断 BERT 输出的表征和 BiLSTM 捕获的长距离语义信息的重要性,按照重要程度加权结合得到新的表征,F1 值为 96.7%。本文提出的 Att-Aux-BERT-BiLSTM-CRF 模型较 BERT-BiLSTM-CRF 的 F1 值提高了 1.5 个百分点,较 BiLSTM-CRF 的 F1 值提高了 2.6 个百分点。

表 4 不同模型的实验结果对比

Tab. 4 Comparison of experimental results of different models

模型	准确率	召回率	F1 值
BiLSTM-CRF	95.5	92.9	94.1
BERT-BiLSTM-CRF	97.6	93.0	95.2
Aux-BERT-BiLSTM-CRF	98.1	94.1	96.0
Att-Aux-BERT-BiLSTM-CRF	98.7	94.7	96.7

3.3.2 不同注意力值计算函数对模型的影响

本文设计了 3 种不同的注意力值计算函数(皮尔逊相关系数、马氏距离、感知机)。为了选择最佳的注意力值计算函数,设计实验比较了这些函数对于模型性能的影响。在实验中,BERT 层输出作为辅助分类层,BiLSTM 层输出作为主分类器层,两者

作为注意力值计算函数的输入。表 5 给出了 Att-Aux-BERT-BiLSTM-CRF 模型在兽药致病语料上使用不同注意力值函数计算的结果。

表 5 不同注意力值计算函数的模型性能

Tab. 5 Influence of different attention calculation

方法	functions on model		
	准确率	召回率	F1 值
感知机	98.2	94.4	96.2
马氏距离	98.4	94.5	96.4
皮尔逊相关系数	98.7	94.7	96.7

实验表明,感知机效果最差,马氏距离优于感知机,使用皮尔逊相关系数取得的效果最好,获得了最高的 F1 值(96.7%)。原因可能是马氏距离和皮尔逊相关系数是机器学习中简单有效的相似性度量方法。两者都对数据进行了规范化。与他们相比,感知机的结构更为复杂,更容易导致过拟合,优化更为困难。所以最终选择实验结果较好的皮尔逊相关系数作为注意力值计算函数。

3.3.3 不同实体实验结果比较

表 6 列出了 Att-Aux-BERT-BiLSTM-CRF 对兽药致病数据集上各类实体的识别结果。由实验可知,相较于药物和不良反应,动物和摄入方式的准确率、召回率和 F1 值都较高。因为动物和摄入方式种类较少且不存在实体嵌套的现象,所以各项指标都较高。兽药致病 4 类实体在不同模型上实验结果对比见图 3。从图 3 可以看出,Att-Aux-BERT-BiLSTM-CRF 的各项指标都较高。

表 6 Att-Aux-BERT-BiLSTM-CRF 在各类实体上的识别结果

Tab. 6 Att-Aux-BERT-BiLSTM-CRF recognition results on various entities

类别	准确率	召回率	F1 值
动物	99.7	98.6	99.2
药物	98.6	94.3	96.3
不良反应	97.1	94.6	95.7
摄入方式	99.6	97.6	98.6

3.3.4 案例分析

最后,本文在 pubchem 数据库中获得兽药致病文献摘要,使用本文提出的 Att-Aux-BERT-BiLSTM-CRF 模型识别兽药致病文献中的实体,结果如表 7 所示。

4 结束语

针对兽药致病领域的命名实体识别任务,提出了基于 Att-Aux-BERT-BiLSTM-CRF 神经网络

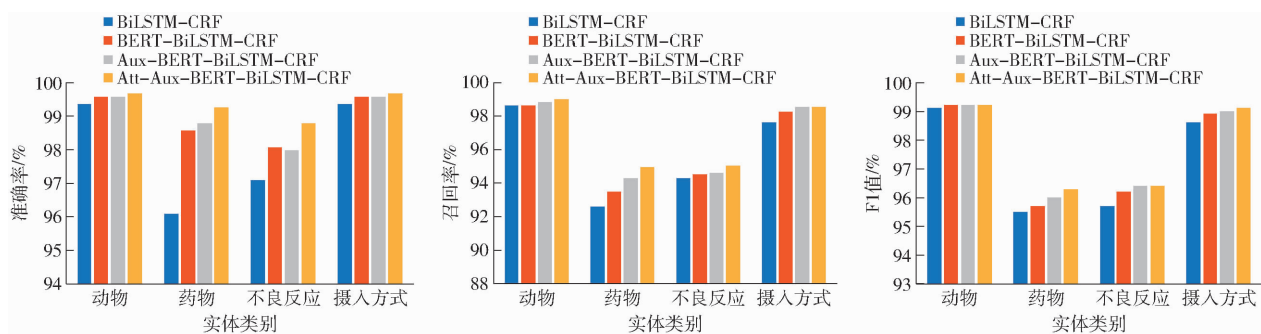


图 3 兽药致命实体实验结果对比

Fig. 3 Comparison of experimental results of animal drugs

表 7 实体识别结果

Tab.7 Entity recognition results

序号	摘要内容	识别结果
1	研究了马拉硫磷对免疫反应的影响	药物: 马拉硫磷
2	将马拉硫磷用于培养的大鼠为期 14 d	药物: 马拉硫磷 动物: 大鼠
3	每天每只给予 2 mg/kg 药剂的动物出现厌食、震颤和共济失调	不良反应: 厌食 不良反应: 震颤 不良反应: 共济失调

模型的方法。通过实验比较了 3 种不同注意力值计算函数(皮尔逊相关系数、马氏距离、感知机)对于模型的影响,最终选择了实验结果较好的皮尔逊相关系数作为注意力值计算函数进行计算。该模型将 BERT 层输出结果作为辅助分类器,BiLSTM 层输出结果作为主分类器,通过注意力机制将两者动态加权融合在一起,丰富了语义信息。与其他典型模型相比,获得了最优的 F1 值(96.7%),说明了其在兽药致命命名实体识别方面的优越性。

参 考 文 献

[1] LUST E B, BARTHOLD C, MALESKER M A, et al. Human health hazards of veterinary medications: information for emergency departments[J]. The Journal of Emergency Medicine,2011, 40(2): 198 - 207.

[2] CAPLETON A, COURAGE C, RUMSBY P, et al. Prioritising veterinary medicines according to their potential indirect human exposure and toxicity profile[J]. Toxicology Letters,2006, 163(3): 213 - 223.

[3] 郑丽敏,齐珊珊,田立军,等. 面向食品安全事件新闻文本的实体关系抽取研究[J]. 农业机械学报,2020,51(7): 244 - 253. ZHENG Limin, QI Shanshan, TIAN Lijun, et al. Entity relation extraction of news texts for food safety events[J]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(7):244 - 253. (in Chinese)

[4] 俞鸿魁,张华平,刘群,等. 基于层叠隐马尔可夫模型的中文命名实体识别[J]. 通信学报,2006,27(2): 87 - 94. YU Hongkui, ZHANG Huaping, LIU Qun, et al. Chinese named entity identification using cascaded hidden Markov model[J]. Journal of Communication,2006,27(2): 87 - 94. (in Chinese)

[5] 李想,魏小红,贾璐,等. 基于条件随机场的农作物病虫害及农药命名实体识别[J]. 农业机械学报,2017, 48(增刊): 178 - 185. LI Xiang, WEI Xiaohong, JIA Lu, et al. Recognition of crops diseases and pesticides named entities in Chinese based on conditional random fields[J]. Transactions of the Chinese Society for Agricultural Machinery,2017,48(Supp.):178 - 185. (in Chinese)

[6] HABIBI M, WEBER L, NEVES M, et al. Deep learning with word embeddings improves biomedical named entity recognition [J]. Bioinformatics,2017, 33(14): 37 - 48.

[7] 孙娟娟,于红,冯艳红,等. 基于深度学习的渔业领域命名实体识别[J]. 大连海洋大学学报,2018,33(2):265 - 269. SUN Juanjuan, YU Hong, FENG Yanhong, et al. Recognition of nominated fishery domain entity based on deep learning architectures[J]. Journal of Dalian Fisheries University,2018,33(2):265 - 269. (in Chinese)

[8] 郭旭超,唐詹,刁磊,等. 基于部首嵌入和注意力机制的病虫草害命名实体识别[J]. 农业机械学报,2020,51(增刊 2): 335 - 343. GUO Xuchao, TANG Zhan, DIAO Lei, et al. Recognition of Chinese agricultural diseases and pests named entity with joint radical-embedding and self-attention mechanism[J]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(Supp. 2):335 - 343. (in Chinese)

[9] 杨锦锋,关毅,何彬,等. 中文电子病历命名实体和实体关系语料库构建[J]. 软件学报,2016, 27(11): 2725 - 2746. YANG Jinfeng, GUAN Yi, HE Bin, et al. Corpus construction for named entities and entity relations on Chinese electronic medical records[J]. Journal of Software,2016, 27(11): 2725 - 2746. (in Chinese)

[10] 李丽双,郭元凯. 基于 CNN - BLSTM - CRF 模型的生物医学命名实体识别[J]. 中文信息学报,2018, 32(1): 116 - 122. LI Lishuang, GUO Yuankai. Biomedical named entity recognition with CNN - BLSTM - CRF [J]. Journal of Chinese Information Processing, 2018,32(1):116 - 122. (in Chinese)

- [11] 鄂海红, 张文静, 肖思琪, 等. 深度学习实体关系抽取研究综述[J]. 软件学报, 2019, 30(6): 1793 - 1818.
E Haihong, ZHANG Wenjing, XIAO Siqu, et al. Survey of entity relationship extraction based on deep learning[J]. Journal of Software, 2019, 30(6): 1793 - 1818. (in Chinese)
- [12] ZENG D, SUN C, LIN L, et al. LSTM - CRF for drug-named entity recognition[J]. Entropy, 2017, 19(6): 283.
- [13] HUANG Z, XU W, YU K. Bidirectional LSTM - CRF models for sequence tagging[J]. Computer Science, 2015, 4(1): 1508 - 1519.
- [14] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171 - 4186.
- [15] 杨飘, 董文永. 基于 BERT 嵌入的中文命名实体识别方法[J]. 计算机工程, 2020, 46(4): 40 - 52.
YANG Piao, DONG Wenyong. Chinese named entity recognition method based on BERT embedding [J]. Computer Engineering & Science, 2020, 46(4): 40 - 52. (in Chinese)
- [16] 谢腾, 杨俊安, 刘辉. 基于 BERT - BiLSTM - CRF 模型的中文实体识别[J]. 计算机系统应用, 2020, 29(7): 48 - 55.
XIE Teng, YANG Jun'an, LIU Hui. Chinese entity recognition based on BERT - BiLSTM - CRF model [J]. Application of Computer System, 2020, 29(7): 48 - 55. (in Chinese)
- [17] SZEGEDY C, WEI L, YANGQING J, et al. Going deeper with convolutions [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1 - 9.
- [18] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 3431 - 3440.
- [19] 张晗, 郭渊博, 李涛. 结合 GAN 与 BiLSTM - Attention - CRF 的领域命名实体识别[J]. 计算机研究与发展, 2019, 56(9): 1851 - 1858.
ZHANG Han, GUO Yuanbo, LI Tao. Domain named entity recognition combining GAN and BiLSTM - Attention - CRF [J]. Journal of Computer Research and Development, 2019, 56(9): 1851 - 1858. (in Chinese)
- [20] REI M, CRICHTON G K O, PYYSALO S. Attending to characters in neural sequence labeling models [EB/OL]. COLING. <https://arxiv.org/pdf/1611.04361>, 2016.
- [21] 罗凌. 生物医学文本挖掘若干关键技术研究[D]. 大连: 大连理工大学, 2019.
- [22] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C] // Advances in Neural Information Processing System, 2013: 3111 - 3119.
- [23] PETERS M, NEUMANN M, IYYER M, et al. Deep contextualized word representations [C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018: 2227 - 2237.
- [24] RADFORD A, NARASIMHAN K, SAKIMANS T, et al. Improving language understanding by generative pre-training [EB/OL]. OpenAi. <https://www.cs.ubc.ca/amuham01/LING530/paper/radford2018improving.pdf>, 2018.
- [25] WANG Q, ZHOU Y M, TONG R, et al. Incorporating dictionaries into deep neural networks for the chinese clinical named entity recognition [J]. Journal of Biomedical Informatics, 2019, 92: 103 - 133.
- [26] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning Research, 2014, 15(1): 1929 - 1958.
- [27] 买买提·阿依甫, 吾守尔·斯拉木, 帕丽旦·木合塔尔, 等. 基于 BiLSTM - CNN - CRF 模型的维吾尔文命名实体识别 [J]. 计算机工程, 2018, 44(8): 230 - 236.
AYIFU Maimaiti, SILAMU Wushouer, MUHETAER Palidan, et al. Uyghur named entity recognition based on BiLSTM - CNN - CRF model [J]. Computer Engineering, 2018, 44(8): 230 - 236. (in Chinese)
- [28] GRIDACH M. Character-level neural network for biomedical named entity recognition [J]. Journal of Biomedical Informatics, 2017, 70: 85 - 91.