文章编号：2096-1472(2022)-08-01-05
DOI:10.19644/j.cnki.issn2096-1472.2022.008.001

基于机器学习的药品不良反应实体识别研究综述

仲雨乐，马诗雯，陆豪杰，韩 普

(南京邮电大学管理学院，江苏 南京 210003)
✉1715272757@qq.com; 1535328066@qq.com; 1071879399@qq.com; hanpu@njupt.edu.cn

摘 要：命名实体识别是信息抽取和知识图谱构建等任务的关键技术环节。近年来，药品不良反应(Adverse Drug Reaction，ADR)实体识别受到学界的极大关注。为了系统梳理ADR实体识别的研究进展及趋势，首先总结了ADR实体识别时常见的问题；其次将ADR实体识别研究划分为基于规则和词典、基于传统机器学习、基于深度学习和基于迁移学习方法四类并进行分析和对比，得出目前基于迁移学习方法的ADR实体识别性能较优；接着给出了ADR实体识别涉及的数据集以及评价指标；最后对ADR实体识别未来研究方向进行了展望。

关键词：自然语言处理；实体识别；深度学习；神经网络模型；药品不良反应
中图分类号：TP391.1 **文献标识码**：A

Survey of Research on Entity Recognition of Adverse Drug Reaction based on Machine Learning

ZHONG Yule, MA Shiwen, LU Haojie, HAN Pu

(*School of Management*, *Nanjing University of Posts and Telecommunications*, *Nanjing* 210003, *China*)
✉1715272757@qq.com; 1535328066@qq.com; 1071879399@qq.com; hanpu@njupt.edu.cn

Abstract: Named entity recognition is a key technical link in tasks such as information extraction and knowledge graph construction. In recent years, entity recognition in adverse drug reaction (ADR) has attracted great attention in academic circles. In order to systematically sort out the research progress and trend of entity recognition in ADR, this paper proposes to summarize the common problems in entity recognition of ADR. Then, the ADR entity identification study is divided into four methods respectively based on rules and dictionary, traditional machine learning, deep learning and transfer learning models. After analyzing and comparing the four methods, it is concluded that the current method based on transfer learning has better ADR entity recognition performance. In addition, the datasets and evaluation indexes involved in ADR entity recognition are given. Finally, the future research directions of entity recognition in ADR are prospected.

Keywords: natural language processing; entity recognition; deep learning; neural network model; adverse drug reaction

www.rjgczz.com

1 引言(Introduction)

药品不良反应(Adverse Drug Reaction，ADR)是指在疾病的预防、诊断、治疗或调节生理功能过程中，患者接受正常剂量的药物时出现的所有有害的，与用药目的无关的反应[1]。药品在使用过程中出现的不良反应往往会对药物的治疗效果产生不好的影响，更为严重的可能会直接危害病人的生命安全，因此药物安全监管与评价体系的重要工作内容是识别、收集及分析ADR[2]。

命名实体识别(Named Entity Recognition，NER)是自然语言处理(Natural Language Processing，NLP)中重要的研究领域，主要是将具有特定意义的实体从非结构化的自由文本中抽取出来。与通用领域的实体不同，ADR中的命名实体类型通常包括药品、适应症、不良反应症状、身体部位等。实体识别是ADR中的关键环节，对ADR知识图谱构建[3]、药物重定位[4]、智能医疗[5]和药物知识发现[6]有着重要意义。

针对ADR实体识别任务，本文系统梳理了国内外近些年来有关ADR实体识别领域的相关文献成果，整理归纳了相关研究的主要方法及研究思路，分析其中的关键技术和不足之处，并给出了未来可能的研究方向和发展趋势。

2 ADR实体识别面临的问题(Problems faced by ADR entity recognition)

尽管实体识别受到医疗健康信息抽取领域学者的极大关

基金项目：国家级大学生创新训练计划项目(SZDG2021040).

注，但在ADR实体识别中仍然面临许多亟待解决的问题，其中在中文领域的ADR文本实体识别方面尤为突出。一方面，中文普遍存在实体嵌套的问题，如“头脑胀痛”既包含身体部位的实体“头脑”，又包含症状类的实体“头脑胀痛”，但目前ADR实体识别常忽略嵌套实体问题，而是选择专注于最外层实体[7]；另一方面，针对英文的生物医学文献[8]、临床记录[9]、社交媒体[10]，ADR实体识别数据集比较完善，国内ADR数据集相对匮乏。由于涉及隐私等问题，学界常用的数据是从社交媒体中爬取的药品不良反应评论，这些数据需要自行人工标注，代价高昂，并且这类数据往往存在大量口语化表述，导致标注的实体不够准确，标注数据规模也非常有限，无法进行大规模的模型训练。

3 ADR中实体识别方法与模型(Entity recognition methods and models in ADR)

根据所采用的研究方法，ADR实体识别方法可分为基于规则和词典的方法、以条件随机场(Conditional Random Field，CRF)为代表的传统机器学习方法、以循环神经网络(Recurrent Neural Network，RNN)与BERT(Bidirectional Encoder Representations from Transformers)为代表的深度学习方法和以对抗性迁移学习为代表的迁移学习方法。

3.1 基于规则和词典的ADR实体识别

实体识别中最早使用的是基于规则和词典的方法，通过文本规则、知识库和专业领域词典实现实体识别。冯变玲等[11]利用关联规则抽取ADR实体并通过计算支持度和可信度来得到药物不良反应。YANG等[12]通过爬虫技术从MedHelp上获取大量用户关于使用药品而产生不良反应的留言信息，利用CHV医学字典进行ADR实体识别，准确识别出选定药品的不良反应，掌握了药物和各自不良反应之间相互联系的内在关系。类似地，SARKER等[13]收集了Twitter和DailyStrenth中与选定药物名称相关的用户评论，以药学词典为基础，完成了评论文本中不良反应的实体识别。

通过对已有研究的分析可知，基于规则和词典的ADR实体识别方法简单易懂，准确率高。但仅仅基于规则和词典适用性较差，无法对ADR文本中较复杂的实体类型进行很好的识别处理，只能识别一些特定类型的实体。因此，基于规则和词典的方法大多时候是为了补充机器学习的方法，用于辅助ADR实体识别模型。

3.2 基于传统机器学习模型的ADR实体识别

随着机器学习不断更新迭代，相关机器学习模型逐渐应用于ADR实体识别领域中。常用的模型有隐马尔可夫模型(Hidden Markov Models，HMM)[14]和条件随机场[15]等。

3.2.1 基于隐马尔可夫模型的ADR实体识别

HMM是ADR实体识别任务的常用模型之一。SAMPATHKUMAR等[16]提出了一个以HMM为基础的文本挖掘系统，从医学论坛Medications中爬取有关ADR的文本数据，经人工标注后作为ADR实体识别的数据集，对手动注释的数据集进行10倍交叉验证发现，HMM的F1值平均为0.76，远高于基线模型。

3.2.2 基于条件随机场的ADR实体识别

CRF[17-18]是一种用于序列标记任务的概率统计模型，在ADR实体识别领域中广泛使用。CRF克服了HMM的独立性假设条件，能够考虑连续标签之间的依赖关系，常与其他模型结合生成全局最优序列。TANG等[19]将CRF和长短期记忆神经网络相结合，在ADR公开语料库CADEC[20]上开展系列实验，实验表明该方法F1值高于CRF。CHRISTOPOULOU等[21]将CRF模型与双向长短期记忆神经网络进行结合，构建ADR实体识别模型，并获得了较好的实验结果。王姝元[22]提出了一种带有CRF的双向长短期记忆体的特定体系结构，将5,865篇Twitter帖子标注为“Drug”“Indication”和“ADR”三种标签来进行实体识别，实验效果明显优于CRF模型以及以词典为核心的实体识别方法。

尽管机器学习推动ADR实体识别任务取得了巨大进展，但在模型实施前期，需要大量时间对数据集进行预处理工作，且机器学习模型难以处理大规模数据，无法使用复杂特征，使得基于机器学习模型的ADR实体识别结果难以进一步提高。

3.3 基于深度学习模型的ADR实体识别

随着时代的发展，深度学习模型在越来越多的领域有着出色的表现，各类自然语言处理任务也逐渐引入该模型[23]。目前实体识别任务以深度学习为核心内容，基于深度学习的实体识别方法非常适用于解决序列标注问题，ADR实体识别由此进入新阶段。

3.3.1 基于循环神经网络的ADR实体识别

循环神经网络[24]是一类以序列数据为输入，在序列的演进方向进行递归且所有节点(循环单元)按链式连接的递归神经网络。PANDEY等[25]搭建了一个带有注意力机制的双向递归神经网络，将双向RNN用于自建数据集编码，然后将其输入生成注意权重的注意力层，并在RNN中使用注意力权重来进行ADR实体识别。

长短期记忆神经网络(Long Short-Term Memory，LSTM)[26]是一种较为特殊的RNN，主要包含遗忘门、输入门、输出门与一个记忆单元。传统的RNN在训练中很容易出现梯度爆炸或者梯度消失的问题，导致无法处理较长的序列数据，而LSTM能够很好地洞悉并获取在字段中相隔较远的依赖关系。双向长短期记忆神经网络(Bidirectional Long Short-Term Memory，BiLSTM)是LSTM的变体，能够更加有效地提取文本的上下文信息，在ADR实体识别中取得了较好的效果。GUPTA等[27]提出了一种基于半监督BiLSTM的ADR实体识别模型，在2007—2010年收集的ADR推文上开展实验，实验F1值比CRF高出3.01%。为了进一步提升模型效果，不少研究往往将BiLSTM与CRF进行结合。朱笑笑等[28]构建了基于BiLSTM与CRF的ADR抽取模型，在Twitter数据集上进行实验，结果表明结合BiLSTM和CRF能够有效识别出社交媒体上不规范的不良反应实体。张亚飞[29]等构建了BiLSTM-CRF实体识别方法，从“39健康药物论坛”上爬取26种药物用户评论数据来识别ADR实体，从而挖掘出药物论坛上潜在的ADR。CHEN等[30]在国内ADR报告上对比了CRF、BiLSTM-CRF和基于词汇特征的BiLSTM-CRF模型实体识别任务的效果，结果表明基于词汇特征的BiLSTM-CRF模型实验F1达到94.35%，表现最优。

GRU(Gated Recurrent Unit，门控循环单元)模型[31]是

www.rjgczz.com

LSTM模型的创新模型之一，其特点在于同时具备LSTM的遗忘、更新机制，且简化了LSTM模型架构，提高了运行速度，降低了梯度弥散的风险。陈瑶等[32]以双向门控循环单元(Bi-directional Gated Recurrent Unit，Bi-GRU)结构为基础，引入注意力机制以及字向量与分词向量优化模型，对国内ADR报告中的ADR过程描述部分进行“药品-不良反应”的实体识别和关系抽取研究，取得了较好效果。

3.3.2 基于混合神经网络的ADR实体识别

为了解决单一特征的神经网络模型不能充分利用文本信息的问题，赵鑫等[33]提出了一种混合多特征的神经网络模型，使用包含卷积计算且具有深度结构的卷积神经网络(Convolutional Neural Networks，CNN)[34]和LSTM对ADR文本进行实体识别研究，研究表明混合模型效果优于单一特征的神经网络模型。杨文明等[35]以BiLSTM-CRF为基准模型，设计了IndRNN-CRF和IDCNN-BiLSTM-CRF两种不同的实体识别模型，并从寻医问药等网站爬取有关药物不良反应的评论构建数据集，在此数据集上验证模型的有效性。对比实验结果可知，IDCNN-BiLSTM-CRF模型在精确率、召回率和F1值上均超过了基准模型BiLSTM-CRF模型，总体性能最好。TIFTIKCI等[36]集成了BiLSTM、CNN和CRF用于ADR实体识别，并构建了基于规则的方法规范已识别的ADR实体，在2017 年文本分析会议药物不良反应挑战测试数据集上进行评估，结果表明由深度学习构建的实体识别架构和基于规则的实体规范化模型组成的系统是从药物标签中提取ADR的一种很有前途的方法。

3.3.3 基于预训练模型的ADR实体识别

传统的实体识别方法将字映射成One-hot编码，难以表明一词多义。为了解决这一问题，学者们提出了一种利用预训练语言模型进行单词表示的方法[7]。虽然该方法计算较为复杂，但可以充分捕捉长距离依赖关系，在ADR实体识别任务中常用来生成词向量。

早期运用较为广泛的预训练模型大多以Word2Vec[37]工具对词向量进行训练，LIN等[38]采用Word2Vec单词嵌入方法识别Twitter数据集上的ADR，实验结果表明Word2Vec在缩小特征空间的同时让召回率和总体F1值都达到更高。GUPTA等[27]使用Word2Vec在大型通用语料库上训练词向量作为ADR实体识别模型的输入。然而Word2Vec仅关注到了词的部分信息，没有考虑到词与局部窗口外词的联系，所以后续出现了GloVe[39]和ELMo[40]算法。GloVe利用共现矩阵，同时考虑了局部信息和整体信息[39]。ELMo则针对Word2Vec和GloVe存在词在不同的语境下有不同的含义，而向量表示却相同这一问题进行了优化，使其能够学习到单词用法的复杂特性。PANDEY等[25]采用Word2Vec和GloVe从大量的ADR资源中实现了无监督的单词嵌入来改善RNN，提高了ADR实体识别的准确性。

随着预训练模型受到越来越多的关注，更多的预训练模型用于ADR实体识别。2018 年，DEVLIN等[41]提出了BERT模型，其关键部分是Transformer结构，它是一个基于自注意力机制的深度网络[42]。朱晓旭[43]在ADR的实体识别阶段，以BiLSTM-CRF模型为基础，融入基于医学文献的预训练语言模型Bio-BERT作为词表示，并加入自注意力机制来更好地处理长距离的依赖问题。WU等[44]将BERT、BiLSTM与CRF相结合，令微调BERT模型获得的词特征作为输入，通过BiLSTM-CRF进行特征提取，准确识别出国内ADR报告中的目标实体，远优于其他基线模型。

从目前的研究成果来看，深度学习模型是端到端的训练方式，自动提取特征，能够学习ADR文本当中的深层语义信息，在各种ADR实体识别任务中均取得较好结果。近年来，学者们尝试在基于神经网络的结构上引入注意力机制、迁移学习、图神经网络、远程监督学习等以进一步提升ADR实体识别效果。

3.4 基于迁移学习方法的ADR实体识别

迁移学习是深度学习前沿的研究领域，许多学者利用深度学习技术构建迁移学习模型，将源域和模型的相关信息迁移到目标域，提高目标领域的实验效果，能够有效解决ADR实体识别任务资源匮乏、实体类型多样化等问题，提升了ADR实体识别效果。

ZHANG等[45]提出了一种针对ADR实体识别任务的对抗性迁移学习模型，引入共享的字符级卷积神经网络(charCNN)学习两个域数据集之间相似的字符级特征，通过融合来自PubMed(源域)的生物医学领域信息，提高Twitter数据集(目标域)上的ADR实体识别效果，并根据两个域的比例设置最终损失函数平衡任务。该方法的F1值在Twitter数据集上为68.58%，实现了最优的性能。KANG等[46]构建了由对抗性迁移学习组成的神经网络模型ASAIBC，以分词数据集为源域，以人工标注的新疆地区ADR数据集为目标域，结合自注意力机制和独立循环神经网络(Independently Recurrent Neural Network，IndRNN)捕捉句子整体依赖性，充分利用中文分词任务和NER任务的边界来实现信息共享，该模型F1评分均值为90.57%，优于现有模型，可以进一步提高ADR实体识别效果。

4 数据集和评估效果(Datasets and evaluating performance)

4.1 ADR实体识别常用数据集

在ADR实体识别研究早期，常见的ADR数据来源有自发报告系统、电子健康记录(EHR)、医学文献等，其中自发报告系统所收集的ADR报告和电子健康记录较为常用。但是这些数据来源依旧有很多弊端，自发报告系统就存在错报漏报、上报缺乏主观能动性的缺陷，EHR文档所涵盖的数据范畴也并不全面，因此，系统全面的数据来源是开展ADR实体识别研究的关键。

随着互联网的盛行，社交媒体已经成为近年来ADR数据量大幅增长的领域之一。基于社交网络的评论给ADR实体识别研究提供了宝贵的数据源，除了Twitter和微博社交平台，还有一些专注于疾病和健康领域的网站，如国外的Dailystrenth、MedHelp、SteadyHealth以及国内的“39健康网”“丁香园”和“寻医问药”，吸引一大批用户通过发帖方式表达用药或治疗经历，或是向线上专业的医护人员进行提问。这些包含大量ADR信息的网站近期成为ADR研究的重要数据源。除此以外，一些公开的语料数据库或者比赛提供的数据集也是ADR实体识别研究常用的数据集，如MIMIC-III[47]语料数据库和n2c2数据集常用于各类实体识别研究。

www.rjgcz.com

由此可见，无论是自发报告系统、社交媒体或是公开数据库，都为ADR实体识别研究提供了新的视野，使得近几年来ADR实体识别研究取得了明显进展。

4.2 数据集标注策略

ADR实体识别数据集一般采用BIO、BIOE和BIOES等标注格式。在BIO标注格式中，B表示实体开端字符(Begin)，I表示内部(Inside)，O表示外部(Outside)。BIOES是BIO的扩展，在BIO的基础上，E表示一个词处于某个实体的结束，S表示这个词自己就可以组成一个实体(Single)，O含义不变。在标注数据时，常采用以上两种标注策略来进行ADR实体识别。由于BIOES具有更强的指向性和可识别度，因此能显著提高识别效率[48]，但后续出现的深度学习模型中学者们会更倾向于使用BIO标注。

ADR数据集通常有13种标签结果，分别为："O""B-Drug（药品）""I-Drug（药品）""B-Dose（剂量）""I-Dose（剂量）""B-Indication（适应症）""I-Indication（适应症）""B-Part（部位）""I-Part（部位）""B-Symptom（症状）""I-Symptom（症状）""B-Time（时间）""I-Time（时间）"。如"效果还行但副作用好大，吃了几天后头晕嗜睡，四肢无力"，经BIO策略标注后为"效/O果/O还/O行/O但/O副/O作/O用/O好/O大/O，/O吃/O了/O几/B-Time天/I-Time后/O头/B-Symptom晕/I-Symptom嗜/I-Symptom睡/I-Symptom，/O四/B-Part肢/I-Part无/B-Symptom力/I-Symptom。/O"。

4.3 评价指标

ADR实体识别任务一般采用的评价指标包括精确率(Precision)、召回率(Recall)及F1值(F1-Measure)等。具体计算公式如下：

$$P=\frac{TP}{TP+FP} \tag{1}$$

$$R=\frac{TP}{TP+FN} \tag{2}$$

$$F1=\frac{2\times P\times R}{P+R} \tag{3}$$

其中，*TP*表示原本属于正例的样本被准确地分类为正例的数量，即正确识别为ADR实体的数量；*FP*表示原本为负例的样本被错误地分类为正例的数量，即错误识别为ADR个体的数量；*FN*表示原本属于正例的样本被错误地分类为负例的数量，即错误地识别为非不良反应的数量。

5 ADR实体识别未来研究趋势(Future research trend of ADR entity recognition)

近年来科技发展迅速，各种实体识别方法层出不穷，但是如何解决ADR文本的复杂性、数据稀缺和实体嵌套等疑难点，更好地开展ADR实体识别研究仍然面临着诸多挑战。基于ADR实体识别广阔的应用前景，未来可以从以下三个方面对ADR实体识别开展研究：

(1)模型创新，引入混合深度学习、持续学习、主动学习和多任务学习等前沿学习方法。混合深度学习将传统机器学习与主流的深度学习模型相结合，充分利用各学习方法的优势提高对多义词、低频实体等复杂实体的识别性能；持续学习在多个时间段的数据流中学习，建立知识记忆以便在新知识的学习过程中不会忘记旧有的特征，减轻灾难性遗忘；主动学习通过选择最有用未标记样本，降低人工标注成本，有利于解决实体识别数据标注问题；多任务学习同时学习多个相关任务，相比于单任务学习，多任务学习能够共享任务信息，提升泛化效果。这些前沿学习方法目前已经常用于生物医学命名实体识别领域，且效果较基础模型有显著提高[49-52]，在该领域发挥了较为可观的作用，因此后续可以在已有研究的基础上引入这些学习方法来进一步提高ADR实体识别效果。

(2)多特征融合，融合部首、词性、字词、词典、依存句法分析及外部语义等多重特征。近年来，多特征融合在医疗实体识别领域应用广泛[53-55]，能够多方面抽取医学专业领域知识，准确且完整地提取句子的特征信息，增强词与词之间的关联性，充分挖掘文本上下文深层语义信息。而目前ADR实体识别领域相关模型向量特征表示过于单一，多特征融合方法尚未充分应用，因此未来融合多特征也是提升ADR实体识别性能的一个方案。

(3)构建中文领域ADR实体识别大规模高质量数据集。目前ADR实体识别研究通常需要大量的标注数据，然而标注庞大的数据集需要大量的人力、时间和更多的领域专家知识。因此，如何利用小规模的标注语料构建大规模的数据集是一个可参考的方向[56]。未来或许可以整合现存的多个语料库，利用自动化的方法进行数据集构建来提高ADR实体识别的准确性。

6 结论(Conclusion)

目前，ADR实体识别是药物警戒和智能医疗的热门研究领域。本文首先分析了ADR实体识别目前面临的问题；其次将现有ADR实体识别领域的研究方法进行归纳和分类，总结了包括基于规则和词典、传统机器学习、深度学习和迁移学习方法这四类ADR实体识别的方法与模型，分析得到当前基于迁移学习的方法在ADR实体识别任务上性能较优；接着介绍了ADR实体识别常用的数据集和当前较为流行的数据集标注策略以及评估指标；最后给出了ADR实体识别研究的研究趋势，预计未来可以通过模型创新、多特征融合和构建大规模高质量数据集来提升ADR实体识别效果，为该领域深入研究提供参考。

参考文献(References)

[1] 朱敏,王艳.63例药物不良反应报告分析[J].现代医药卫生,2008(03):448-449.

[2] EDWARDS I R, ARONSON J K. Adverse drug reactions: definitions, diagnosis, and management[J]. The Lancet, 2000, 356(9237):1255-1259.

[3] 魏伟,权勤波,王玲华,等.基于战略坐标分析的我国核心期刊药品不良反应领域知识图谱研究[J].药物流行病学杂志,2020,29(10):714-720.

[4] 张永祥,程肖蕊,周文霞.药物重定位——网络药理学的重要应用领域[J].中国药理学与毒理学杂志,2012,26(06):779-786.

[5] 钟文艳.美国智能医疗产业发展现状分析[J].全球科技经济瞭望,2017,32(06):38-44.

[6] 范馨月,崔雷.基于文本挖掘的药物副作用知识发现研究[J].数据分析与知识发现,2018,2(03):79-86.

[7] 王红,王彩雨.中文医疗命名实体识别方法研究综述[J].山东

www.rjgcs.com

师范大学学报(自然科学版),2021,36(02):109-117.
[8] KANG N, SINGH B, BUI C, et al. Knowledge-based extraction of adverse drug events from biomedical text[J]. BMC Bioinformatics, 2014, 15(1):1-8.
[9] HENRIKSSON A, KVIST M, DALIANIS H, et al. Identifying adverse drug event information in clinical notes with distributional semantic representations of context[J]. Journal of Biomedical Informatics, 2015, 57:333-349.
[10] LIU J, ZHAO S, ZHANG X. An ensemble method for extracting adverse drug events from social media[J]. Artificial Intelligence in Medicine, 2016, 70:62-76.
[11] 冯变玲,杨世民,贺小红,等.药品不良反应多维关联规则挖掘及预警模型构建[J].中国药事,2012,26(10):1076-1082.
[12] YANG C C, YANG H, JIANG L, et al. Social media mining for drug safety signal detection[C]// CHRISTOPHER C. Proceedings of the 2012 International Workshop on Smart Health and Wellbeing. New York: Association for Computing Machinery, 2012:33-40.
[13] SARKER A, GONZALEZ G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training[J]. Journal of Biomedical Informatics, 2015, 53:196-207.
[14] RABINER L, JUANG B. An introduction to hidden Markov models[J]. IEEE ASSP Magazine, 1986, 3(1):4-16.
[15] SUTTON C, MCCALLUM A. An introduction to conditional random fields[J]. Foundations and Trends® in Machine Learning, 2012, 4(4):267-373.
[16] SAMPATHKUMAR H, CHEN X, LUO B. Mining adverse drug reactions from online healthcare forums using hidden Markov model[J]. BMC Medical Informatics and Decision Making, 2014, 14(1):1-18.
[17] LAFFERTY J, MCCALLUM A, PEREIRA F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]// SANTOS E T. Proceedings of 18th International Conference on Machine Learning. São Paulo: Springer, 2001:282-289.
[18] WALLACH H M. Conditional random fields: An introduction[J]. Technical Reports, 2004, 53(2):267-272.
[19] TANG B, HU J, WANG X, et al. Recognizing continuous and discontinuous adverse drug reaction mentions from social media using LSTM-CRF[J/OL]. (2018-04-19) [2022-07-08]. https://doi.org/10.1155/2018/2379208.
[20] KARIMI S, METKE-JIMENEZ A, KEMP M, et al. CADEC: A corpus of adverse drug event annotations[J]. Journal of Biomedical Informatics, 2015, 55:73-81.
[21] CHRISTOPOULOU F, TRAN T T, SAHU S K, et al. Adverse drug events and medication relation extraction in electronic health records with ensemble deep learning methods[J]. Journal of the American Medical Informatics Association, 2020, 27(1):39-46.
[22] 王姝元.基于社交媒体的药物不良反应发现研究与应用[D].大连:大连理工大学,2018.
[23] 韩兆宇.基于神经网络集成的命名实体识别研究[D].徐州:中国矿业大学,2019.
[24] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization[J/OL]. (2015-02-19) [2022-07-08]. https://arxiv.org/abs/1409.2329v5.
[25] PANDEY C, IBRAHIM Z, WU H, et al. Improving RNN with attention and embedding for adverse drug reactions[C]// KOSTKOVA P. Proceedings of the 2017 International Conference on Digital Health. New York: Association for Computing Machinery, 2017:67-71.
[26] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
[27] GUPTA S, PAWAR S, RAMRAKHIYANI N, et al. Semi-supervised recurrent neural network for adverse drug reaction mention extraction[J]. BMC Bioinformatics, 2018, 19(8):1-7.
[28] 朱笑笑,杨尊琦,刘婧.基于Bi-LSTM和CRF的药品不良反应抽取模型构建[J].数据分析与知识发现,2019,3(02):90-97.
[29] 张亚飞,于琦,王于心,等.基于药物论坛中潜在不良反应与适应症的知识发现体系构建[J].中华医学图书情报杂志,2020,29(07):38-43.
[30] CHEN Y, ZHOU C, LI T, et al. Named entity recognition from Chinese adverse drug event reports with lexical feature based BiLSTM-CRF and tri-training[J]. Journal of Biomedical Informatics, 2019, 96:103252.
[31] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J/OL]. (2014-09-03) [2022-07-08]. https://arxiv.org/abs/1406.1078.
[32] 陈瑶,吴红,葛卫红,等.基于深度学习模型的我国药品不良反应报告实体关系抽取研究[J].中国药科大学学报,2019,50(06):753-759.
[33] 赵鑫,李正光,吴镝,等.多特征混合神经网络模型在ADR分类中的应用[J].大连交通大学学报,2018,39(04):116-120.
[34] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]// JORDAN M I. Advances in Neural Information Processing Systems. Red Hook: Curran Associates, 2012:1097-1105.
[35] 杨文明,褚伟杰.在线医疗问答文本的命名实体识别[J].计算机系统应用,2019,28(02):8-14.
[36] TIFTIKCI M, ÖZGÜR A, HE Y, et al. Machine learning-based identification and rule-based normalization of adverse drug reactions in drug labels[J]. BMC Bioinformatics, 2019, 20(21):1-9.
[37] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]// BURGES C J C. Proceedings of NIPS. Red Hook: Curran Associates, 2013:3111-3119.
[38] LIN W S, DAI H J, JONNAGADDALA J, et al. Utilizing different word representation methods for twitter data in adverse drug reactions extraction [C]//PENG Y H.

www.rigcz.com

Proceedings of the 2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI). Hong Kong, China: IEEE, 2015:260-265.

[39] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]// MOSCHITTI A. Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. Washington: IEEE Press, 2014:1532-1543.

[40] ILIC S, MARRESE-TAYLOR E, BALAZS J A, et al. Deep contextualized word representations for detecting sarcasm and irony[C]// BALAHUR A. Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Brussels: SSMA, 2018:2-7.

[41] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[J/OL]. (2019-05-24) [2022-07-08]. https://arxiv.org/abs/1810.04805v1.

[42] 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(07):48-55.

[43] 朱晓旭.面向社交媒体的药物关系挖掘研究[D].大连:大连理工大学,2020.

[44] WU H, JI J, TIAN H, et al. Chinese-named entity recognition from adverse drug event records: Radical embedding-combined dynamic embedding-based bert in a bidirectional long short-term conditional random field (Bi-LSTM-CRF) model[J]. JMIR Medical Informatics, 2021, 9(12):e26407.

[45] ZHANG T, LIN H, REN Y, et al. Identifying adverse drug reaction entities from social media with adversarial transfer learning model[J]. Neurocomputing, 2021, 453:254-262.

[46] KANG K, TIAN S, YU L. Named entity recognition of local adverse drug reactions in Xinjiang based on transfer learning[J]. Journal of Intelligent & Fuzzy Systems, 2021, 40(5):8899-8914.

[47] JOHNSON A E W, POLLARD T J, SHEN L, et al. MIMIC-III, a freely accessible critical care database[J]. Scientific Data, 2016, 3(1):1-9.

[48] YANG J, LIANG S, ZHANG Y. Design challenges and misconceptions in neural sequence labeling[C]// ZHAO D. Proceedings of the 27th International Conference on Computational Linguistics. Gyeongju: ICCL, 2018:3879-3889.

[49] LIU J, GAO L, GUO S, et al. A hybrid deep-learning approach for complex biochemical named entity recognition[J]. Knowledge-Based Systems, 2021, 221:106958.

[50] LEE C S, LEE A Y. Clinical applications of continual learning machine learning[J]. The Lancet Digital Health, 2020, 2(6):e279-e281.

[51] NASEEM U, KHUSHI M, KHAN S K, et al. A comparative analysis of active learning for biomedical text mining[J]. Applied System Innovation, 2021, 4(1):23.

[52] MEHMOOD T, GERVINI A E, LAVELLI A, et al. Combining multi-task learning with transfer learning for biomedical named entity recognition[J]. Procedia Computer Science, 2020, 176:848-857.

[53] 李丹,徐童,郑毅,等.部首感知的中文医疗命名实体识别[J].中文信息学报,2020,34(12):54-64.

[54] 许力,李建华.基于句法依存分析的图网络生物医学命名实体识别[J].计算机应用,2021,41(02):357-362.

[55] 巩敦卫,张永凯,郭一楠,等.融合多特征嵌入与注意力机制的中文电子病历命名实体识别[J].工程科学学报,2021,43(09):1190-1196.

[56] 郑强,刘齐军,王正华,等.生物医学命名实体识别的研究与进展[J].计算机应用研究,2010,27(03):811-815,832.

作者简介：

仲雨乐(2001-)，男，本科生.研究领域：自然语言处理.

马诗雯(2001-)，女，本科生.研究领域：信息抽取.

陆豪杰(2001-)，男，本科生.研究领域：数据挖掘.

韩　普(1983-)，男，博士，副教授.研究领域：数据分析.

(上接第33页)

[3] 李延臣,车德勇,于银成,等.基于SolidWorks的静电除尘器结构参数化设计研究[J].电力建设,2011,32(05):80-83.

[4] 刘康奇,仲梁维.基于SQL Server与SolidWorks二次开发的型钢库的设计[J].农业装备与车辆工程,2020,58(05):115-118,122.

[5] 刘玲.基于嵌入式电源远程监控系统[D].南京:南京理工大学,2008.

[6] 田继涛,黄晓华,张言中,等.基于SolidWorks的盾构机刀具参数化设计[J].煤矿机械,2019,40(11):186-188.

[7] 杨慧萍,郑婷,黄雄波.基于OLE技术的国企退休人员档案系统设计[J].无线互联科技,2021,18(13):51-53.

[8] 黄洁.基于ArcGIS组件的电子海图显示系统的研究与实现[D].武汉:武汉理工大学,2008.

[9] 王喆.基于C#.net的铁路供电设备电气试验与保护整定辅助管理软件设计[D].石家庄:石家庄铁道大学,2021.

[10] 曾婧雯,任军学,尹佳,等.TC18钛合金铣削刀具几何参数优化研究[J].航空精密制造技术,2013,49(06):37-40.

[11] ARIZMENDI M, CAMPA F J, FERNANDEZ J, et al. Model for surface topography prediction in peripheral milling considering tool vibration[J]. CIRP Annals-Manufacturing Technology, 2009, 58(1):93-96.

作者简介：

崔建昆(1964-)，男，博士，副教授.研究领域：齿轮传动，车辆工程.

丁佳乐(1997-)，男，硕士生.研究领域：机械设计，软件开发.

夏　娟(1997-)，女，硕士生.研究领域：机械传动，铣削工艺.

www.rjgczz.com