

中文医学命名实体识别算法研究^{*}

李梦蝶 张 平

李功利 姜 伟

(电子科技大学生命科学与技术学院 成都 610054)

(四川省屏山县人民医院 屏山 645350)

李 科

蔡培强

(电子科技大学生命科学与技术学院 成都 610054)

(四川省屏山县人民医院 屏山 645350)

〔摘要〕 针对中文医学命名实体识别任务中实体细粒度较大、识别准确率不高等问题,提出一种融合特征 Albert 的中文医学命名实体识别算法,利用自建的真实标注语料对模型进行训练与测试,结果表明模型具有较好的识别效果。

〔关键词〕 中文电子病历;命名实体识别算法;融合特征;Albert 算法

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2022.03.009

Study on Chinese Medical Named Entity Recognition Algorithm LI Mengdie, ZHANG Ping, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China; LI Gongli, JIANG Wei, Pingshan People's Hospital of Sichuan Province, Pingshan 645350, China; LI Ke, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China; CAI Peiqiang, Pingshan People's Hospital of Sichuan Province, Pingshan 645350, China

〔Abstract〕 Aiming at the problems of large fine-grained entity and low recognition accuracy in Chinese medical Named Entity Recognition (NER) task, an algorithm of Chinese medical NER based on fusion features and Albert is proposed. The model is trained and tested by self-built real labeled corpus, and the results show that the model has good recognition effect.

〔Keywords〕 Chinese Electronic Medical Records (EMR); Named Entity Recognition (NER) algorithm; fusion features; Albert algorithm

1 引言

1.1 研究背景与意义

电子病历 (Electronic Medical Records, EMR)

是患者就诊过程中生成的电子化病历信息,包括医生问诊产生的文本数据,实验室检查产生的数字、图像、图表数据,这些信息为医疗诊断和医学研究提供重要数据支持^[1]。大量有用信息存在于叙述性文本中,计算机能够自动利用的结构化

〔收稿日期〕 2021-06-01

〔作者简介〕 李梦蝶,硕士研究生;通讯作者:蔡培强,主任医师。

〔基金项目〕 2020 年中央引导地方科技发展专项科技扶贫类项目“屏山县多级联动互联网医院及智能特色专科平台研发与应用”(项目编号:2020ZYD002);四川省科技支撑项目“四川省肿瘤随访登记大数据服务平台建设”(项目编号:2019JDPT0008);四川省科技支撑项目“基于床垫式多参数监测系统的伤残病人智慧医疗监护系统”(项目编号:21ZDYF3601)。

数据有限, 采用自然语言处理技术进行文本结构化成为医学信息学研究热点^[2]。命名实体识别 (Named Entity Recognition, NER) 是自然语言处理中的重要方法, 该技术可识别医学文本中特定的医学实体, 标识其蕴含信息, 加快医院信息化进程, 还可以提供医疗决策支持, 改善医疗质量, 提高医疗人员工作效率, 减少临床决策失误, 促进临床医学研究。

1.2 命名实体识别研究方法

1.2.1 实体识别常用的机器学习方法 基于词典和规则是最早的命名实体识别方法。李毅、保鹏飞和薛万国^[3]基于中文电子病历特点, 构建一套生物医学命名实体规则、实体词和用于分类的词表。实体识别常用的机器学习方法包括隐马尔科夫模型 (Hidden Markov Model, HMM), 支持向量机 (Support Vector Machine, SVM), 条件随机场模型 (Conditional Random Fields, CRF) 等^[4]。Zhang J、Shen D 和 Zhou G 等^[5]提出第 1 个处理级联现象的 HMM 实体识别系统, 该系统不仅集合拼写、形态学多种特征, 还加入命名实体缩写, 在 GENA3.0 语料上 F 值达到 66.5%。Lee K J、Hwang Y S 和 Kim S 等^[6]提出一个基于 SVM 的蛋白质和 DNA 实体识别系统, 该系统在确认已知词性标注的实体边界后, 利用单词信息和上下文构造特征集进行实体识别, 在 GENA3.0 语料上测试 $F1$ 值达到 74.8%。2016 年 Bhasuran B、Murugesan G 和 Abdulkadhar S 等^[7]以 CRF 作为基础分类器, 结合模糊字符串匹配算法识别罕见疾病名称, 该应用在 CDR (Chemical Disease Relations) 语料库的训练集上测试 F 值达到 89.12%。词典与机器学习相结合的方法也常用来进行实体识别^[8]。

1.2.2 神经网络算法 逐渐被迁移到自然语言处理中, 常用来实体识别的一般有卷积神经网络和递归神经网络。Zhao Z、Yang Z 和 Luo L 等^[9]基于多标签的卷积神经网络模型进行疾病识别, 级联了字符、单词、词典特征, 最终模型在美国国家生物技术信息中心 (National Center for Biotech-

nology Information, NCBI) 数据集上测试 F 值达到 85.17%。2019 年 Cho H 和 Lee H^[10]提出一个基于双向长短时记忆 (Bi-Directional Long Short-Term Memory, BiLSTM) 和 CRF 的生物医学命名实体识别系统, 在 NCBI 和 CDR 语料库上都取得不错的识别效果。

1.2.3 预训练模型 由于具备可提供初始化参数等优点而常用于医学实体识别。Zhang X、Zhang Y 和 Zhang Q 等^[11]利用 Bert (Bidirectional Encoder Representation from Transformers) 模型训练词向量, 该模型在其自建的癌症语料上测试 $F1$ 值达到 93.53%。罗凌、杨志豪和宋雅文等^[12]训练一个基于 ELMo (Embeddings from Language Models) 语言模型的命名实体模型, 模型中引入包含汉字内部结构的笔画信息, 最终在中国知识图谱与语义计算大会数据集上测试 F 值达到 91.75%。

1.2.4 融合特征 Albert 的中文医学命名实体识别算法 基于规则和词典识别效果理想, 但建立模板、维护词典成本较高, 且移植性差。机器学习解决了移植性问题, 但其严格特征提取影响实体识别效果。深度学习克服了机器学习中的问题, 但神经网络模型参数多, 医学数据集较小, 容易过拟合。基于通用语料训练的预训练模型提供初始化参数, 在下游任务中微调即可, 具有更好的泛化能力, 在小量的医学实体语料上表现不错。综上, 本文提出一种融合特征 Albert (A Lite Bert, 轻量级 Bert) 的中文医学命名实体识别算法, 该算法旨在训练融合文本短语、语言语义的字向量特征, 将融合的字向量特征与句子向量特征相加以提高实体识别效果。

2 模型结构

2.1 构建融合特征 Albert + BiLSTM + CRF 模型

针对现有研究方法局限性提出一种融合特征 Albert 的中文医学命名实体识别算法。其核心在于训练的字向量融合文本短语、语言学、语义学特征, 然后将融合的字向量特征和句子向量特征相

加, 最终可以更高效地进行命名实体识别。与其他字向量特征模型相比, 本文算法能够充分学习到预测文本的字信息、词语信息、语言语义信息、句子信息、语篇级信息。算法流程实现分为数据预处理、词嵌入、句子嵌入、命名实体识别 4 个阶段, 见图 1。

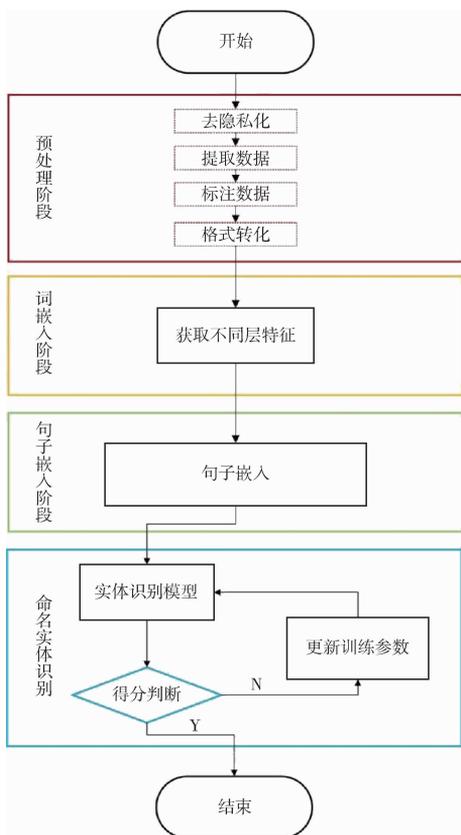


图 1 融合特征 Albert 的中文医学命名实体识别算法结构

2.2 词嵌入阶段

2.2.1 Bert 模型 2018 年 Bert 模型被提出^[13], 旨在解决以往预训练模型不能学习文本上下文特征的问题。模型主要框架是编码器模型, 将完形填空和下一个句子预测作为训练目标, 基于大规模数据训练词向量能有效捕捉文本前后文特征, 见图 2。

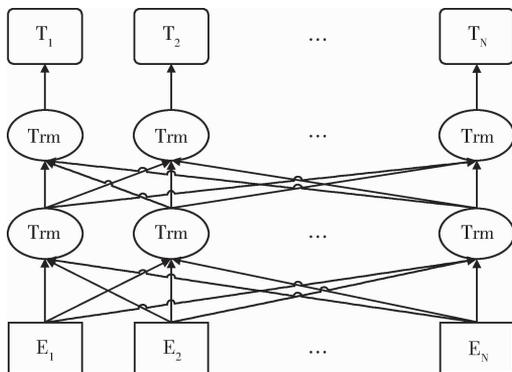


图 2 Bert 模型结构

2.2.2 Albert 模型 主干结构与 Bert 相似, 在 Bert 的基础上主要进行 3 方面改进, 分别是嵌入因式分解参数化、跨层参数共享、句子间一致性损失, 减少了参数但是性能有较大提升^[14]。嵌入因式分解参数化在 Bert 中字嵌入 E 和隐藏层 H 是相等的。语言模型建模时, 字嵌入和隐藏层分离可以更充分地利用模型参数。实际应用上训练词典大小为 V , 嵌入矩阵为 $V * E$, 如果 H 变大 E 随着变大, 此时嵌入矩阵 $V * E$ 将会变大, 最终导致训练过程中数十亿参数仅有少部分更新。而 Albert 中采用的因式分解是先将 one-hot 向量映射到低维空间, 然后映射到隐藏层中, 最终使模型参数从 $O(V * H)$ 减少到 $O(V * E + E * H)$ 。与仅共享前馈神经网络层参数和仅共享注意力层参数不同, Albert 选择共享所有层参数, 通过测量 $L2$ 距离和余弦相似度, 表明跨层共享参数的方法能够稳定网络参数。在 Bert 模型中以预测下一个句子为训练目标时, 属于下一句来自同一文档连续片段, 属于非下一句来自不同文档片段, 这种包含主题预测和一致性预测的方法较为简单。因此 Albert 提出基于句子顺序预测损失, 正例的选取与 Bert 相同, 反例是相同的两个连续片段但顺序交换。本文发现 Albert 算法在低层网络中学习词语特征、中层网络中学习语言学特征、高层中学习语义特征, 所以提出融合不同层的训练特征, 用来解决单个字向量或词语向量问题。

2.3 句子嵌入阶段

2.3.1 概述 长短期记忆 (Long Short - Term Memory, LSTM) 是循环神经网络 (Recurrent Neural Network, RNN) 的改进版, 有效解决了梯度消失和梯度下降问题, 而 BiLSTM 是由前向 LSTM 和后向 LSTM 构成的算法。

2.3.2 LSTM 结构 (图 3) 一个 LSTM 由输入状态 x_t 、细胞状态 C_t 、临时细胞状态 \tilde{C}_t 、隐藏层状态 h_t 、遗忘门 f_t 、记忆门 i_t 、输出门 O_t 构成。从输入词到隐藏层状态序列过程如下: 将上一时刻细胞状态中需要遗忘和记忆的信息, 传递给后续时刻的单元 LSTM, 更新信息后输出当前时刻的隐藏层状态。

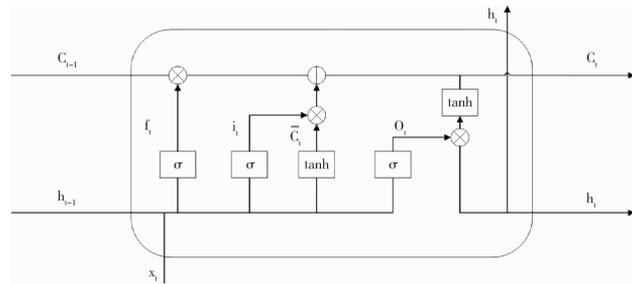


图 3 LSTM 网络结构

2.3.3 LSTM 计算过程 在确定当前时刻的隐藏层状态时, 先计算需要被遗忘的信息。遗忘门输出由当前时刻输入状态和上一时刻隐藏层状态控制, 如公式 (2-1) 所示, 输出范围从 0 到 1。

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

确定丢弃信息后, 计算需要更新的信息, 如公式 (2-2) 所示, 其输出范围也是 0 到 1。临时细胞状态 \tilde{C}_t 是 tanh 层创建的新候选值向量, 如公式 (2-3) 所示, 其输出范围为 -1 到 1。

$$i_t = \sigma(W_i * (h_{t-1}, x_t) + b_i) \quad (2-2)$$

$$\tilde{C}_t = \tanh(W_c * (h_{t-1}, x_t) + b_c) \quad (2-3)$$

丢弃和更新信息确定后, 第 3 步是将上一时刻细胞状态更新为当前时刻细胞状态。如公式 (2-4) 所示。

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (2-4)$$

最后就是单元结构 LSTM 输出。输出门由 sig-

moid 层、当前时刻输入、上一时刻隐藏层状态决定, 如公式 (2-5) 所示。当前时刻输出门状态乘以通过 tanh 层变换的当前时刻细胞状态即可得到当前时刻隐藏层状态输出 h_t , 见式 (2-6)。

$$O_t = \sigma(W_o * (h_{t-1}, x_t) + b_o) \quad (2-5)$$

$$h_t = O_t * \tanh(C_t) \quad (2-6)$$

上列式中 σ 是 sigmoid 函数, h_{t-1} 是上一时刻的隐藏层状态, x_t 是当前时刻输入状态, W_f, W_i, W_c, W_o 是矩阵参数, b_f, b_i, b_c, b_o 是偏差, f_t 为遗忘门输出, i_t 为记忆门输出, \tilde{C}_t 为临时细胞状态, C_t 就是更新后的细胞状态。

2.3.4 BiLSTM 算法 上述 4 步是一个完整单元的 LSTM 计算过程, 将前向的 LSTM 和后向的 LSTM 拼接即本文使用的 BiLSTM。对句子“咳嗽伴头晕”编码, 前向 LSTM 输入“咳嗽”“伴”“头晕”, 输出隐藏层状态 $\{h_{L0}, h_{L1}, h_{L2}\}$, 后向 LSTM 输入“头晕”“伴”“咳嗽”, 输出隐藏层状态 $\{h_{R0}, h_{R1}, h_{R2}\}$, 将前向后向的隐藏层状态拼接得到具有上下文特征的隐藏层状态 $\{[h_{L0}, h_{R0}], [h_{L1}, h_{R1}], [h_{L2}, h_{R2}]\}$, 即 $\{h_0, h_1, h_2\}$ 。拼接后的 h_i 就是输入词对应的预测标签, 来自于实体分类中概率最大的。

2.4 构建命名实体识别模型

为解决 BiLSTM 的输出仅在每一时刻选出一个最大的概率标签, 而各个时刻输出之间不存在关系的问题, 设计引入具有转移矩阵的 CRF, 在其中加入约束规则以保证最终预测结果的合法性。假设给定一个输入序列 X , 其预测标签序列为 y , 输入序列的预测标签概率为 $p(y | X)$ 。

$$X = (X_1, X_2, \dots, X_n) \quad (2-7)$$

$$y = (y_1, y_2, \dots, y_n) \quad (2-8)$$

$$p(y | X) = \frac{e^{s(X,y)}}{\sum_{y \in Y_X} e^{s(X,y)}} \quad (2-9)$$

$$s(X,y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=1}^n P_{i, y_i} \quad (2-10)$$

其中, $A_{y_i, y_{i+1}}$ 为从 y_i 到 y_{i+1} 的转移矩阵, 大小为 $(n+2) * (n+2)$, n 为所有标签个数, $+2$ 是增加了起止位置。 P_{i, y_i} 为第 i 个位置 softmax 输出为 y_i 的概率, 起止位置记为 0。在计算预测得分

$s(X, y)$ 中, 加上实体标签之间转移概率, 给模型加入约束规则: 句子总是以 “B-” 或 “O” 开始, 开头出现 “I-” 时就是错误标签; “I- 实体标签 a ” 出现在 “B- 实体标签 a ” 之前时, 该标签无效; “B- 实体标签 a ” “I- 实体标签 b ” 是非法标签。利用最大似然法训练转移矩阵参数, 其损失函数如公式 (2-11) 所示, 待模型训练完毕, 采用维特比算法找出最优路径, 也就是更合理的输入序列的预测标签。

$$-\log(p(y|X)) = \log\left(\sum_{y \in Y_x} e^{s(X, \bar{y})}\right) - s(X, y) \quad (2-11)$$

最终构建完成本文的医学实体识别算法, 如图 4 所示, 输出预测文本 “咳嗽伴头晕”, 在 Albert 阶段, 模型学习预测文本内的短语、语言语义学特征, 并将融合这些特征的字向量输出, 作为 BiLSTM 输入, 以进一步学习预测文本的上下文特征, 然后在 CRF 层加入实体名称约束规则, 最终可输出较为合理的实体标签预测结果。

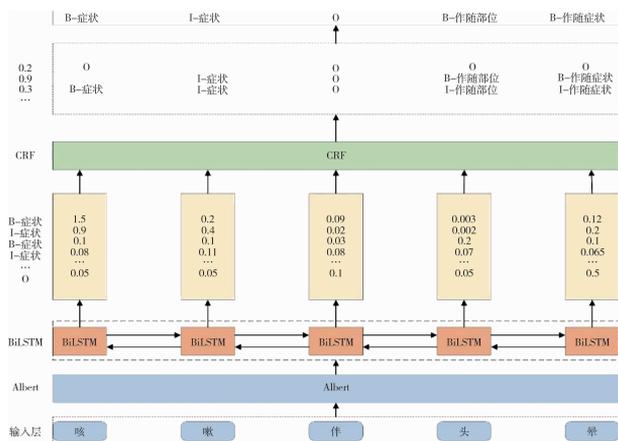


图 4 融合特征的 Albert + BiLSTM + CRF 模型

3 实验及结果分析

3.1 实验环境及评价标准

实验在 Inter core i5 - 4590 3.3GHz 的 CPU 上运行, 能够为文中涉及命名实体识别模型提供硬件支持。所有模型均采用 Python 语言编写, 版本号为 3.6, 用 Tensorflow 搭建神经网络学习框架。使用命名实体识别领域的 3 个指标作为模型衡量标准: 准确率 P (Precision)、召回率 R (Recall) 以及 $F1$ 值

(F -score), TT 表示真实命名实体是正例, 预测结果也是正例, FT 表示真实命名实体是反例, 但预测结果为正例, TF 表示真实命名实体是正例, 但预测结果为反例, FF 表示真实命名实体是反例, 预测结果也为反例。

$$P = \frac{TT}{TT + FT} \quad (3-1)$$

$$R = \frac{TT}{TT + TF} \quad (3-2)$$

$$F1 = \frac{R \times P \times (1 + \beta^2)}{R + P \times \beta^2} \quad (3-3)$$

3.2 实验数据

3.2.1 概述 本研究数据来自临床电子病历并进行脱敏处理。最终完成 900 份辅助检查、200 份体格检查、5 000 份主诉的语料标注, 分别列作 3 个实验项目, 辅助检查项目中实体数量最多的是检查指标、最少的是检查值, 体格检查项目中实体数量最多的是结果、最少的是项目值, 主诉项目中实体数量最多的是症状、最少的是伴随部位。

3.2.2 标注规则及实体类别阶段 在临床医生指导下制定 3 个类目的标注规则, 分别是辅助检查、体格检查、主诉。辅助检查实体包括检查日期、检查项目、检查结果、检查值、检查指标、指标结果、指标值。体格检查实体包括部位、查体项目、结果、状态、项目值。主诉实体包括症状、部位、持续时间、伴随症状、伴随部位、伴随时间、既往史。

3.2.3 数据标注阶段 语料库构建采用 BIO 标注体系, 标注工具是开源 doccano 软件。标注小组由 3 人组成, 各环节有异议的部分讨论确定并立即更新标注规则。

3.3 实验及实验参数

为验证本模型的可行性和有效性设计 3 组对比实验, 对比模型分别是未融合字向量特征的 Albert + BiLSTM + CRF、BiLSTM + CRF 和 Bert + BiLSTM + CRF。本研究词嵌入的最大句子长度为 300, 学习率为 $2e - 05$, 迭代次数为 40, 这些参数已经使模型识别效果达到最佳。

3.4 实验结果

3.4.1 与未融合向量模型的识别效果对比 实验训练集与测试集为 7: 3。提出的融合字向量特征与未融合字向量特征相比, 在主诉这个语料库中表现最明显, 精确率 P 明显高于 2.0%, 召回率 R 明显高于 1.0%, $F1$ 值高于 2.0%。在辅助检查的语料库中表现次之, 精确率和 $F1$ 值明显高于 1.5%, 召回率 R 高于 1.4%。在体格检查语料库上表现略低, 不过 $F1$ 值高于 0.4%, 见表 1。

表 1 本模型与未融合向量模型在 3 个项目上的识别效果对比 (%)

项目	未融合向量模型			本模型		
	P	R	$F1$	P	R	$F1$
辅助检查	94.19	95.58	94.88	95.97	97.00	96.48
体格检查	95.32	96.42	95.87	95.99	96.61	96.30
主诉	84.30	87.32	85.78	86.99	88.64	87.81

实验结果表明本文提出的词嵌入阶段融合字向

量特征可以有效提取字的特征信息, 与单纯词嵌入相比, 提升了命名实体识别成功率, 同时也证明本文提出的融合字向量特征是可行的。

3.4.2 与其他经典模型的识别效果对比 与 BiLSTM + CRF、Bert + BiLSTM + CRF 进行对比可以发现加入 Bert 的 BiLSTM + CRF 模型相比于原始模型 $F1$ 值均提高了 2% 以上。在 BiLSTM + CRF 模型中, BiLSTM 被用来学习文本的句子特征, CRF 用来预测输出最优实体标签, 此模型缺少字向量特征, 而 Bert 可以学习到字符级别信息, 所以加入 Bert 的 BiLSTM + CRF 模型具有更好的识别效果。另外, 本文提出的融合字向量特征模型相比于 Bert + BiLSTM + CRF 模型 $F1$ 值均有所提高, 平均 $F1$ 值达到 93.53%。虽然 Bert 和 Albert 都能学习到字符级别的特征, 但在 Albert 进行字向量训练时高中低层学习到的特征不同, 所以将这些特征融合可以有效提高实体识别能力, 见表 2。综合精确率 P 、召回率 R 、 $F1$ 值 3 个指标, 本文提出的模型效果更好, 具有更佳的能力。

表 2 本模型与其他经典模型在 3 个项目上的识别效果对比 (%)

项目	BiLSTM + CRF			Bert + BiLSTM + CRF			本模型		
	P	R	$F1$	P	R	$F1$	P	R	$F1$
辅助检查	92.34	93.64	92.99	94.60	95.50	95.05	95.97	97.00	96.48
体格检查	92.16	94.29	93.22	94.19	96.56	95.36	95.99	96.61	96.30
主诉	81.53	84.92	83.19	86.11	89.54	87.79	86.99	88.64	87.81

4 结语

本文提出一种融合 Albert 不同层训练特征的中文医学命名实体识别算法。算法结合中文电子病历的语言结构特点, 在词嵌入阶段进行优化, 使之可提取预测文本中短语、语言学、语义学特征, 然后融合这些特征并与句子特征相加进行实体识别。基于临床病历, 利用 BIO 标注体系构建辅助检查、体格检查、主诉 3 个语料, 利用这些语料训练测试本算法, 然后与其他模型进行对比, 结果表明本文提出的融合特征模型在 3 个项目上的 $F1$ 值都高于其

他 3 个模型, 平均 $F1$ 值达到 93.53%, 具有更好的识别效果。本文构建的语料库量还比较小, 对模型识别效果有一定影响, 未来可引入主动学习、众包标注等方法构建更大的语料库; 可加入修饰词的识别, 丰富算法识别广度。

参考文献

- 1 杨兵, 聂铁铮, 申德荣, 等. 一种面向医学文本数据的结构化信息抽取方法 [J]. 小型微型计算机系统, 2019, 40 (7): 1479 - 1485.
- 2 郎君, 秦兵, 刘挺, 等. 中文人称名词短语单复数自动识别 [J]. 自动化学报, 2008 (8): 972 - 979.

- 3 李毅, 保鹏飞, 薛万国. 中文电子病历的信息抽取研究 [J]. 生物医学工程杂志, 2010, 27 (4): 757 - 762.
- 4 潘瑾然, 施维, 薛均, 等. 生物医学命名实体识别研究现状及中文生物医学命名实体识别难点与意义综述 [J]. 医学信息学杂志, 2018, 39 (3): 57 - 63.
- 5 Zhang J, Shen D, Zhou G, et al. Enhancing Hmm - Based Biomedical Named Entity Recognition by Studying Special Phenomena [J]. Journal of Biomedical Informatics, 2004, 37 (6): 411 - 422.
- 6 Lee K J, Hwang Y S, Kim S, et al. Biomedical Named Entity Recognition Using Two - Phase Model Based on Svms [J]. Journal of Biomedical Informatics, 2004, 37 (6): 436 - 447.
- 7 Bhasuran B, Murugesan G, Abdulkadhar S, et al. Stacked Ensemble Combined with Fuzzy Matching for Biomedical Named Entity Recognition of Diseases [J]. Journal of Biomedical Informatics, 2016 (64): 1 - 9.
- 8 夏光辉, 李军莲, 阮学平. 基于实体词典与机器学习的基因命名实体识别 [J]. 医学信息学杂志, 2015, 36 (12): 54 - 60.
- 9 Zhao Z, Yang Z, Luo L, et al. Disease Named Entity Recognition from Biomedical Literature Using a Novel Convolutional Neural Network [J]. BMC Medical Genomics, 2017, 10 (Suppl 5): 73 - 84.
- 10 Cho H, Lee H. Biomedical Named Entity Recognition Using Deep Neural Networks with Contextual Information [J]. BMC Bioinformatics, 2019 (20): 735 - 746.
- 11 Zhang X, Zhang Y, Zhang Q, et al. Extracting Comprehensive Clinical Information for Breast Cancer Using Deep Learning Methods [J]. International Journal of Medical Informatics, 2019 (132): 103985.
- 12 罗凌, 杨志豪, 宋雅文, 等. 基于笔画 elmo 和多任务学习的中文电子病历命名实体识别研究 [J]. 计算机学报, 2020, 43 (10): 1943 - 1957.
- 13 Devlin J, Chang M W, Lee K, et al. Bert: Pre - Training of Deep Bidirectional Transformers for Language Understanding [EB/OL]. [2018 - 11 - 13]. <https://arxiv.org/pdf/1810.04805.pdf>.
- 14 Lan Z, Chen M, Goodman S, et al. Albert: A Lite Bert for Self - Supervised Learning of Language Representations [EB/OL]. [2019 - 09 - 29]. <https://openreview.net/pdf?id=H1eA7AEtVS>.

(上接第 44 页)

参考文献

- 1 郑述谱. 关于术语及术语学 [J]. 外语学刊, 2001 (2): 5 - 9, 112.
- 2 冯志伟. 现代术语学引论 (增订本) [M]. 北京: 商务印书馆, 2011.
- 3 魏向清. 人文社科术语翻译中的术语属性 [J]. 外语学刊, 2010 (6): 165 - 167.
- 4 马清海. 试论科技翻译的标准和科技术语的翻译原则 [J]. 中国翻译, 1997 (1): 28 - 29.
- 5 方小兵. 面向翻译的术语变异研究——以语言学术语为例 [M]. 南京: 东南大学出版社, 2012.
- 6 黄昭厚. 谈谈科技术语的审定与统一 [J]. 中国科技术语, 2010, 12 (1): 18 - 22.
- 7 王希杰. 语言的规范化和言语的得体性 [J]. 语言教学与研究, 1998 (1): 69 - 84.
- 8 毛群安, 李长宁, 宋军. 现代互联网与健康类资讯传播 [J]. 中国健康教育, 2017, 33 (7): 579 - 580, 586.
- 9 MacLean D L, Heer J. Identifying Medical Terms in Patient - authored Text: A Crowdsourcing - based Approach [J]. Journal of the American Medical Informatics Association, 2013, 20 (6): 1120 - 1127.
- 10 Chen J, Druhl E, Ramesh B P, et al. A Natural Language Processing System That Links Medical Terms in Electronic Health Record Notes to Lay Definitions: System Development Using Physician Reviews [J]. Journal of Medical Internet Research, 2018, 20 (1): e26.
- 11 全国科技名词委. 全国科学技术名词审定委员会科学技术名词审定原则及方法 [EB/OL]. [2021 - 03 - 27]. <http://www.cnctst.cn/sdgb/sdyzjff/>.
- 12 李曙光. 医学术语翻译中的伦理问题——以 autism 的汉译为例 [J]. 外语研究, 2017, 34 (1): 75 - 79.