

文章编号: 1003-0077(2021)04-0044-07

# 一种文本相似度与 BERT 模型融合的手术操作术语归一化方法

杨飞洪, 孙海霞, 李 姣

(中国医学科学院/北京协和医学院 医学信息研究所, 北京 100020)

**摘 要:** 该文探究手术操作术语归一化方法的构建。首先, 分析手术操作术语归一化数据集的特点; 其次, 调研术语归一化的相关方法; 最后, 结合调研知悉的技术理论方法和数据集特征, 建立手术操作术语归一化模型。该文融合文本相似度排序+BERT 模型匹配开展建模, 在 2019 年中文健康信息处理会议(CHIP2019)手术操作术语归一化学术评测中, 验证集准确率为 88.35%, 测试集准确率为 88.51%, 在所有参赛队伍中排名第 5。

**关键词:** 手术术语; 归一化; BERT; 文本相似度

**中图分类号:** R319

**文献标识码:** A

## A Method for Surgery Term Normalization Based on Text Similarity and BERT Model

YANG Feihong, SUN Haixia, LI Jiao

(Institute of Medical Information, Chinese Academy of Medical Sciences/Peking Union Medical College,  
Beijing 100020, China)

**Abstract:** To explore the method for surgery term normalization, this paper proposes a method of combining text similarity and BERT model. The model scheme is the text similarity ranking + BERT sentence pair matching model. This paper also analyzes the characteristics of the normalized surgery terms, and provides the related methods of clinical term normalization. In the CHIP2019 surgical term normalization task, the accuracy of this method on the verification set is 88.35%, and the accuracy on the test set is 88.51%, and the system based on this method ranked 5th among all participating teams.

**Keywords:** surgery terms; normalization; BERT; text similarity

## 0 引言

手术操作原始词是指医生书写的手术操作名称。医生个人书写习惯的差异会导致手术操作原始词与编码标准词之间存在细微差异, 如使用手术操作的英文缩写等。这些细微差异对手术操作规范和管理不利。国际疾病分类体系(ICD)于 1975 年开始对手术操作进行编码规范, 经过数十年的发展, 目前较为成熟的中文手术操作编码方案有: 北京协和医院依据 ICD-9 手术操作编码体系结合临床实践编制的《ICD9-2017 协和临床版》, 目前在各医院广泛使用的《手术操作分类代码国家临床版 1.1》以及正在逐步推行的《手术操作分类代码国家临床版

2.0》<sup>[1]</sup>。各编码体系的核心是手术操作代码和相应的手术操作标准术语, 如表 1 所示。

表 1 手术操作编码示例

编码	手术操作术语
00.0101	头部血管治疗性超声
00.7301	人工股骨头修复术

统一的手术操作编码体系有利于施行医院绩效评估、医保结算以及各级医联体间的患者转诊。因而, 手术操作编码体系更新导致的历史文本数据, 以及医生的书写习惯造成的手术操作不规范用语, 需要利用计算机技术进行文本数据归一化处理。

国外自 ShARe/CLEF eHealth 2013 发布标注

收稿日期: 2020-07-21 定稿日期: 2020-09-08

**基金项目:** 中国医学科学院医学与健康科技创新工程(2018-I2M-AI-016); 中国医学科学院中央级公益性科研院所基本科研业务费(2018PT33024)

的临床术语归一化语料伊始, SemEval-2014 Task 7 和 SemEval-2015 Task 14 相继发布了英文的临床术语归一化语料。在这些语料上, 研究者们纷纷构建了各种临床术语归一化模型, 并取得不少研究成果。第五届中国健康信息处理会议(CHIP2019)临床术语归一化任务(简称 Task1)<sup>[2]</sup>开放了手术操作归一化数据集, 其中训练集包含手术操作原始词与标准词共计 4 000 对, 验证集 1 000 对, 测试集 2 000 条(未标注), 各数据集术语分布基本一致。Task1 数据集标注依据的手术操作编码体系为《ICD9-2017 协和临床版》。评测任务具体内容是给定一个手术操作原始词, 预测其在编码体系中对应的标准词。Task1 采用准确率评估模型效果, 如式(1)所示。

$$\text{准确率} = \frac{\text{预测准确的手术原始词与标准词数}}{\text{需要预测的手术原始词总数}} \quad (1)$$

本文首先分析 Task1 数据集样本的特点, 对样本数据进行清洗; 其次, 调研临床术语归一化的相关技术方法; 最后, 结合数据集的特点和相关技术方法, 构建临床手术操作原始词归一化模型。

## 1 相关工作

本节从共指消解、临床术语归一化和 ShARe/CLEF 2013 评测任务三个方面开展文献综述。

### 1.1 共指消解

从某种角度上看, Task1 其实是手术操作术语的共指消解问题, 即根据手术操作术语的各类表述寻找共同指代的手术操作编码标准词。

宋洋等人<sup>[3]</sup>研究了基于机器学习的中文共指消解方法。他们依据国际著名评测 ACE (automatic content extraction, 自动内容抽取) 对共指消解的定义, 共指消解即为文本中的表述(mention, 或称为指称语), 确定其在真实世界中所指向的实体(entity)的过程。总结了有监督的机器学习方法处理共指消解时的三个关键步骤: ①表述或实体的检测; ②机器学习模型选择; ③训练数据的构建。

表述检测(也被称为指称语检测, 英文即 mention detection), 即识别出文本中可能产生共指的全部候选表述(候选表述来自于全部的名词短语集合的某个子集), 是共指消解的第一步。表述检测的一般方法包括基于规则的方法<sup>[4]</sup>、基于学习的方法<sup>[5]</sup>以及规则和统计相结合的方法<sup>[6]</sup>。

机器学习模型主要用于针对表述对或表述—实体对的二元分类任务, 即判断表述—表述是否等价或者表述—实体是否配对。文本分类是自然语言处理领域的经典问题, 可以实现二元分类任务的经典机器学习模型有支持向量机(SVM)<sup>[7]</sup>、文本卷积神经网络(TextCNN)<sup>[8]</sup>、双向长短时记忆网络(Bi-LSTM)<sup>[9]</sup>和 BERT<sup>[10]</sup>等。

机器学习模型训练集的构建, 即从训练集中采样用于构建模型学习的数据集。Mccarthy 等人<sup>[11]</sup>将任何两个不在同一实体(或被称为共指链)中的表述构成负例, 任何两个位于同一实体中的表述构成正例。该方法产生的训练实例数量巨大, 而且负样本数远远大于正样本数, 严重影响机器学习算法的效率, 因此后来很少被采用。Soon 等人<sup>[12]</sup>将每一个表述  $m_i$  与其前面最近的一个共指的表述  $m_j$  ( $j < i$ ) 构成正例, 其间的每个表述都和  $m_i$  构成负例, 能较好地解决过采样问题。

### 1.2 临床术语归一化

Xu D 等人<sup>[13]</sup>将术语归一化的方法总结为两个大类, 其一是基于分类的方法, 其二是基于排序学习的方法。他们认为基于分类的方法有两个缺点, 即需要预测的类别空间大和类别间样本不均衡问题。基于排序学习的方法可分为点排序学习、对排序学习和列表排序学习三种不同形式。最终, 他们构建了一套基于 BERT 模型的表述候选生成器+列表排序学习的术语归一化方法, 在多个生物医学的英文文本数据集中取得了最先进的结果。

### 1.3 相关评测任务

ShARe/CLEF 2013 评测任务 1 开放了 298 份匿名化的临床记录, 其中包含出院小结、心电图报告、超声检查报告等。298 份临床记录中, 199 份是训练集, 共包含原始词与标准词对 4 000 余对; 99 份是测试集, 共包含原始词与标准词对 3 600 余对。

基于 ShARe/CLEF 2013 评测任务 1 开放的数据集, Li 等人<sup>[14]</sup>首先采用基于规则的方法生成候选集合, 然后利用文本卷积神经网络模型(TextCNN)<sup>[8]</sup>对候选集合进行排序, 取排序最靠前的词为归一化的标准词。该模型在测试集上的准确率为 90.3%。Luo 等人<sup>[15]</sup>在总结前人研究的基础上, 提出了一种将传统字典查找与双向长短时记忆网络模型(Bi-LSTM)<sup>[9]</sup>相结合的临床术语归一化方法, 达到了 90.6% 的准确率, 相较于单纯的文本相似度方法

的 88.0% 准确率,提升了 2.6%。Ji 等人<sup>[16]</sup>微调开源的 BERT<sup>[10]</sup>、BioBERT<sup>[17]</sup> 和 ClinicalBERT<sup>[18]</sup> 三个模型,利用这三个模型分别计算原始词与标准词间的相似度评分,在测试集上的准确率达到 91.10%。

## 2 方法

本节从 Task1 数据集的样本特点、数据清洗和 Task1 归一化方案的构建这三个层面展开讨论。

表 2 Task1 临床手术术语原始词及标准词示例

原始词	标准词	关系	占比/%
横结肠造口还纳术	横结肠造口闭合术	1:1	95.025
左甲状腺切除术+右甲状腺部分切除术	单侧甲状腺切除伴他叶部分切除术	n:1	3.25
右尺神经松解前移术	尺神经松解术##尺神经移位术	1:n	1.65
右眼白内障超声乳化抽吸术+人工晶体置入术	置入人工晶状体##白内障晶状体乳化和抽吸	n:n	0.075
根治性全膀胱切除+回肠膀胱术+阑尾切除术	小肠段分离术##根治性膀胱切除术##可控回肠膀胱术##阑尾切除术	m:n	

### 2.2 数据清洗

除了手术操作原始词与标准词之间的关系,我们还观察到部分手术原始词中包含标准词的手术代码,如手术原始词“腮腺肿瘤(肿块)切除术+面神经解剖术((26.2910+04.0701))”中的“26.2910”。这些手术代码由医生手工填写,与《ICD9—2017 协和临床版》手术操作编码中的手术操作代码并不一致。因此,需要利用正则表达式方法将这些原始词中的手术代码剔除。数据清洗后的手术原始词示例如表 3 所示。

表 3 手术原始词数据清洗示例

原始词	数据清洗
腮腺肿瘤(肿块)切除术+面神经解剖术((26.2910+04.0701))	腮腺肿瘤(肿块)切除术+面神经解剖术

Task1 验证集样本特点与训练集基本一致,数据清洗采用同样方法。

### 2.3 手术操作术语归一化方案

观察 Task1 数据集,可以发现手术操作原始词与标准词之间的文本相似度很高,同时手术操作原始词之间相似性也很高,因此可以依据文本相似度

### 2.1 样本特点

分析 Task1 训练集样本,发现多个手术操作原始词之间用“+”分隔,多个标准词之间用“##”分隔。每一条样本中,手术操作原始词与标准词之间的关系有 5 种:1:1 关系占总样本 95.025%,n:1 关系占 3.25%,1:n 关系占 1.65%,n:n 关系和 m:n 关系共占 0.075%。Task1 手术操作原始词与标准词示例,如表 2 所示。

特征排序产生手术操作原始词-标准词对的候选集,也就是共指消解中的表述检测。本文选择 BERT 作为手术操作原始词-标准词句对的二元分类模型。训练集的构建策略为:将训练集每条样本的手术操作原始词与标准词作为正例,将与该手术原始词文本相似度相近而标准词不同的前 5 个手术操作原始词与该样本的标准词组成负例。正负样本比例为 1:5,训练集采样示例如表 4 所示。

表 4 二元分类模型采样示例

原始词	标准词	配对
治疗性超声	治疗性超声	是
心脏治疗性超声	治疗性超声	否
其他治疗性超声	治疗性超声	否
头部血管治疗性超声	治疗性超声	否
颈部血管治疗性超声	治疗性超声	否
周围血管治疗性超声	治疗性超声	否

Task1 手术原始词归一化的整个流程如图 1 所示,可简要概括为:首先依据训练集和手术操作编码表建立标准词-手术操作原始词的映射表,依据映射表进行正负采样用于训练 BERT 二分类模型;预测时,计算验证集或测试集中的手术操作原始词与映射表中手术操作原始词的文本相似度,

依据文本相似度排序取前 5 个映射样本。然后，利用 BERT 模型判断前 5 个映射对中待预测的手术操作原始词与前 5 个映射对的手术操作标准词是否等价。若等价，则对应的标准词作为最终答案返回；不等价，则取排序最靠前的手术操作标准

词返回。

其中，文本相似度的计算，采用最长公共子序列算法，也就是两段文本中相同的连续字符串长度占比越大，两文本相似度越高。该算法在 Python 的 difflib 库中提供了最佳实现。

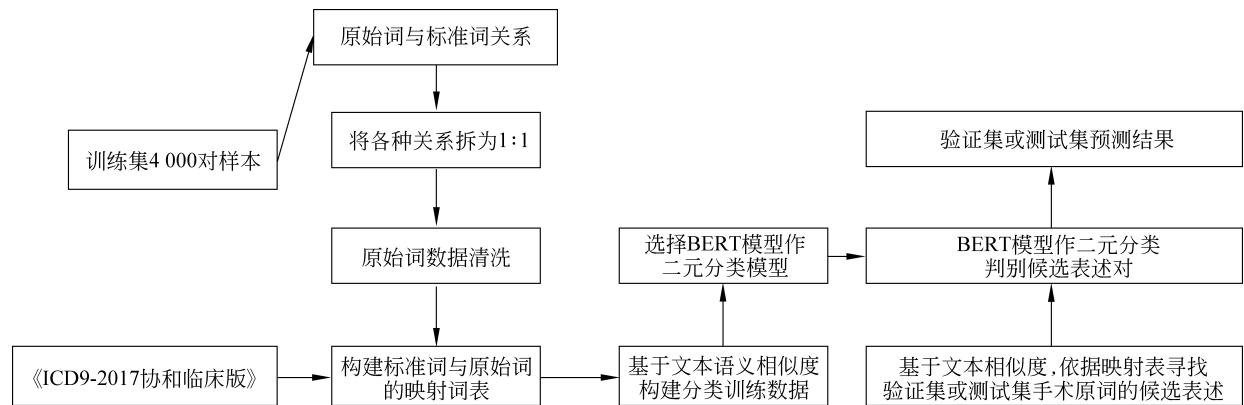


图 1 Task1 手术原始词归一化的流程

训练时，将手术操作原始词和标准词作为句对，和等价标签 1 或不等价标签 0 一起输入 BERT 模型进行训练。预测时，将手术操作原始词和候选标准词作为句对，输入 BERT 模型，预测各句对的标签：为 1 等价，为 0 不等价。同时，BERT 模型可输出相应标签的概率值，返回目标标准词前，可计算预测标签为 0 的概率值与标签为 1 的概率值的差值，将该差值与设定的 BERT 模型参与度阈值( $\alpha$ )做比较，差值大于  $\alpha$  值时才会依据 BERT 模型的结果进行处理，否则返回相似度最高的标准词。

$\alpha$  阈值决定了 BERT 模型在预测过程中的参与度。如  $\alpha=0$  时，表明完全信任 BERT 模型对手术操作原始词和候选标准词的预测结果； $\alpha=0.5$  时，则表明只有当 BERT 模型预测为 0 的概率与预测为 1 的概率之差高于 0.5 时，才继续判断下一个候选句对的标签值，也就是说 BERT 模型在整个标准词的预测中，参与度仅有一半； $\alpha=1$ ，则表明完全不信任 BERT 模型的预测结果，仅使用文本相似度的排序结果返回。标准词预测过程的算法描述，如表 5 所示。

表 5 标准词预测过程算法表述

输入	1 个待预测的手术原始词+5 个候选标准词，进入迭代
1	1 个待预测的手术原始词+1 个候选标准词
2	BERT 模型预测句对的标签
3	标签为 1，则返回；标签为 0，进入步骤 4

续表

4	计算标签为 0 与标签为 1 的概率值之差
5	概率差值 $\geq\alpha$ ，则返回步骤 1，重新寻找标准词；否则，返回当前的候选标准词作为答案
6	若迭代了 5 次，步骤 5 都无返回，则返回最初的标准词，也就是排序最前的标准词
输出	1 个待预测的原始词+1 个预测的标准词

### 3 结果

从 Task1 的训练集中采样了手术操作原始词和标准词的句对正负例样本共计 69 300 余条，采样的正负例样本如表 4 所示。BERT 模型的超参数：最大序列长度(max\_sequence\_len)为 128，批训练子集的大小(train\_batch\_size)为 128，训练周期(epochs)为 3。BERT 二元分类模型在采样的验证集上准确率达到了 98.01%。

预测时，按照 BERT 模型对手术原始词表述与标准词的候选词对筛选的贡献程度，分为 BERT 模型完全参与筛选( $\alpha=0$ )，BERT 模型部分参与筛选( $\alpha=0.5$ )和 BERT 模型完全不参与筛选( $\alpha=1$ )，也就是直接取与待预测手术原始词相似度最高的手术原始词对应的标准词返回)。三组模型在 Task1 验证集上的准确率如表 6 所示，可知 BERT 模型的部分参与筛选可得到最高的准确率。



表 6 三种筛选策略在验证集上的准确率

候选词对筛选策略	准确率/%
BERT 模型完全参与( $\alpha=0$ )	87.35
BERT 模型部分参与( $\alpha=0.5$ )	<b>88.35</b>
BERT 模型完全不参与( $\alpha=1$ )	78.07

按照 BERT 模型部分参与表述对筛选的预测策略,最终在 Task1 测试集上的准确率为 88.51%。

#### 4 消融实验及对比实验

相关研究<sup>[13]</sup>表明,文本检索方法也可以用于获取给定手术原始词的候选标准词。根据文献调研,本文选择 Whoosh<sup>[19]</sup>作为文本检索工具,用于替代最大序列长度算法,实验对比如表 7 所示。Whoosh-based 的方法均没有基于文本相似度的方法准确率高。

表 7 不同方案的归一化准确率( $\alpha=0.5$ )

方法名	准确率/%
最大序列长度算法	78.07
最大序列长度算法+BERT	<b>88.35</b>
最大序列长度算法+LSTM	59.00
最大序列长度算法+TextCNN	23.70
Whoosh-based	69.60
Whoosh-based + BERT	<b>80.50</b>
Whoosh-based + LSTM	29.10
Whoosh-based + TextCNN	29.60

同时,本文也在相同实验环境下实现了基于其他深度学习模型的二分类方法,如 LSTM 和 TextCNN。实验表明,LSTM 和 TextCNN 的表现不如 BERT 模型。

此外,对本文建构的文本相似度+BERT 模型方法进行消融实验。实验表明,“最大序列长度算法+BERT”方法的验证集准确率为 88.35%,相比仅使用“最大序列长度算法”提升了 10.28%。

CHIP2020 发布了新型的临床术语归一化任务,相较于 CHIP2019 的手术术语归一化任务,CHIP2020 包含了更多的临床术语类型,训练集有 8 000 个临床术语+标准词的词对。验证集及测试集还未公开。本文为验证方法的可扩展性,将 CHIP2020 的训练集以 8:2 的比例切分为新的训

练集与测试集后,本文所述方法的效率表现如表 8 所示。实验表明,基于文本相似度的方法[也就是, BERT 模型完全不参与( $\alpha=1$ )]的准确率为 25.69%,这说明该数据集中临床原始词与标准词之间的文本相似度较低,然而本文提出的“文本相似度+BERT 模型”方法准确率为 31.88%,依然可以提升 6.19%。

表 8 三种筛选策略在 CHIP2020 中的准确率

候选词对筛选策略	准确率/%
BERT 模型完全参与( $\alpha=0$ )	31.00
BERT 模型部分参与( $\alpha=0.5$ )	<b>31.88</b>
BERT 模型完全不参与( $\alpha=1$ )	25.69

#### 5 讨论

深度学习模型在手术原始词候选词对筛选中的参与度阈值的设置,需要进一步探讨。我们尝试将 BERT 模型的参与度阈值设为  $\alpha=0.8$ ,验证集上的准确率为 86.80%。可见参与度阈值的设置对模型准确率有直接影响。

手术原始词的表述检测方案需要进一步研究。本文选择文本相似度作为手术原始词表述检测,导致预测结果中出现文本重心偏移的情况出现,如验证集中手术原始词“腹膜外子宫下段剖宫产术”的正确标准词为“腹膜外剖宫产”,而模型预测结果为“低位子宫下段剖宫产”。添加规则的表述检测方法也许能够避免类似情况的出现,进而提高模型预测准确率。

从预测结果文件选取代表性的正确与错误样本各 3 例,进行模型预测结果分析如表 9 所示。观察正确样例可知:①手术原始词与标准词的文本相似度越高,越容易被正确预测,如原始词“右尺神经松解术”和标准词“尺神经松解术”;②待测手术原始词与训练集中的手术原始词相似度越高,越容易被正确预测,如待测手术原始词“右肺上叶后段切除术”,虽然该词与标准词“肺节段切除术”相似度不高,但是与该标准词的映射原始词“右肺下叶背段切除术”有较高的相似度;③针对手术原始词的拆分处理可以较好预测  $n:n$  的数据类型,如手术原始词“T12 椎体活检+椎体成形术”经“+”拆分后,可以被模型正确预测为标准词“椎骨活组织检查##经皮椎骨成形术”。观察错误样例可知:①模型对于原始词:

标准词关系为  $1:n$  或  $m:n$  的样本预测不全;②模型没有考虑原始词经拆分后发生上下丢失的情况,如原始词“腹腔镜右侧多囊肾去顶减压+被膜剥脱术”被拆分为“腹腔镜右侧多囊肾去顶减压”和“被膜剥脱术”,导致“被膜剥脱术”上下文缺失,从而被错误预测为“筋膜剥脱术”。

表 9 正确与错误样例

手术原始词	标准词	预测词	正/误
右尺神经松解术	尺神经松解术	尺神经松解术	✓
右肺上叶后段切除术	肺节段切除术	肺节段切除术	✓
T12 椎体活检+椎体成形术	椎骨活组织检查##经皮椎骨成形术	椎骨活组织检查##经皮椎骨成形术	✓
膀胱镜碎石取石术	经尿道膀胱取石术##经尿道膀胱清除术	经尿道膀胱清除术	×
腹腔镜右侧多囊肾去顶减压+被膜剥脱术	肾包膜剥脱术##腹腔镜下肾囊肿去顶术	腹腔镜下肾囊肿去顶术##筋膜剥脱术	×
腹腔镜下根治性膀胱切除术+原位回肠膀胱术	腹腔镜下膀胱根治切除术##小肠段分离术##可控回肠膀胱术	腹腔镜下膀胱根治切除术##可控回肠膀胱术	×

综上所述,未来可对本文所提架构模型进行如下三个方面的改进:①有针对性地采用表述候选方法,本文所提文本相似度方法在 CHIP2019 的手术术语归一化数据集中效果优异,然而用于 CHIP2020 的临床术语归一化任务时表现不佳,这表明候选表述的提取方法需要因地制宜;②应对原始词与标准词关系为  $1:n$  或  $m:n$  时,可以考虑利用深度学习模型先进行原始词与标准词的关系预测;③原始词的候选表述召回率越高越好,召回率越高说明标准词出现在候选表述中的概率越大,深度学习模型可以发现标准词的概率也越高。

## 6 结论

本文首先探究了 CHIP 临床术语归一化任务(Task1)训练集及验证集的样本特点并进行了数据清洗工作;其次,将 Task1 视作共指消解任务,从文本共指消解角度讨论了基于学习模型的共指消解的关键步骤;最后,依据共指消解的三个关键步骤,结合 Task1 样本数据的特点,构建了基于相似度的手

术操作原始词排序+BERT 二元分类的预测模型,该模型在 Task1 验证集上获得了最高 88.35% 的准确率,在测试集上获得了 88.51% 的准确率。

此外,本文讨论了可以改进模型准确率的三个思考方向,即二元分类模型的选择、学习模型参与度阈值的设置以及手术原始词表述检测方案的制定。这三个方向对模型准确率的提升是否有帮助,值得后续进一步研究。

## 参考文献

- [1] 医政医管局. 关于启动 2019 年全国三级公立医院绩效考核有关工作的通知[EB/OL]. <http://www.nhc.gov.cn/yzygj/s3593g/201904/b8323261bb8a4175a2046d2-fffa93936.shtml>. [2019-11-14].
- [2] 第五届中国健康信息处理会议(CHIP) 评测 1: 临床术语归一化任务[EB/OL]. <http://cips-chip.org.cn/evaluation>. [2019-11-06].
- [3] 宋洋, 王厚峰. 共指消解研究方法综述[J]. 中文信息学报, 2015, 29(1): 1-12.
- [4] Uryupina O, Saha S, Ekbal A, et al. Multi-metric optimization for coreference: The UniTN/IITP/Essex submission to the 2011 CoNLL shared task[C]//Proceedings of the 15th Conference on Computational Natural Language Learning: Shared Task, 2011; Portland, Oregon: Association for Computational Linguistics; 2011: 61-65.
- [5] Zhou G, Kong F. Global learning of noun phrase anaphoricity in coreference resolution via label propagation[C]//Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Singapore; 2009: 978-986.
- [6] 孔芳, 朱巧明, 周国栋. 中英文指代消解中待消解项识别的研究[J]. 计算机研究与发展, 2012, 49(05): 1072-1085.
- [7] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [8] Kim Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014: 1746-1751.
- [9] Liu P, Qiu X, Huang X. Recurrent neural network for text classification with multi-task learning[C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence, New York, USA, 2016: 2873-2879.
- [10] Devlin J, Chang M, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Confer-

- ence of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota, 2019:4171-4186.
- [11] McCarthy J F, Lehner W G. Using decision trees for coreference resolution[C]//Proceedings of the International Joint Conferences on Artificial Intelligence, New York, USA, 1995:1050-1055.
- [12] Soon W M, Ng H T, Lim C Y. A machine learning approach to coreference resolution of noun phrases[J]. Computational Linguistics, 2001, 27: 521-544.
- [13] Xu D, Zhang Z, Bethard S. A generate-and-rank for framework with semantic type regularization for biomedical concept normalization[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020.
- [14] Li H, Chen Q, Tang B, et al. CNN-based ranking for biomedical entity normalization[J]. BMC Bioinformatics, 2017, 18(Suppl 11): 385.
- [15] Luo YF, Sun W, Rumshisky A. A hybrid normalization method for medical concepts in clinical narrative using semantic matching.[C]//Proceedings of AMIA Joint Summits on Translational Science, 2019: 732-740.
- [16] Ji Z, Wei Q, Xu H. BERT-based ranking for biomedical entity normalization[J]. arXiv, preprint arXiv: 1908.03548, 2019.
- [17] Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining[J]. Bioinformatics, 2019, 36 (4): 1234-1240.
- [18] Huang K, Altosaar J, Rangnath R. ClinicalBERT: Modeling clinical notes and predicting hospital readmission[J]. arXiv, preprint arXiv:1904.05342, 2019.
- [19] Chaput M. Whoosh is a fast, pure Python search engine library[CP/OL]. [2020-8-21] <https://whoosh.readthedocs.io/en/latest/intro.html>.



杨飞洪(1995—), 硕士研究生, 主要研究领域为医学自然语言处理。

E-mail: yang.feihong@imicams.ac.cn



李姣(1981—), 通信作者, 博士, 研究员, 主要研究领域为医学信息学, 医学人工智能。

E-mail: li.jiao@imicams.ac.cn



孙海霞(1985—), 博士, 副研究员, 主要研究领域为医学知识图谱, 医学知识表示。

E-mail: sun.haixia@imicams.ac.cn