DOI:10.12138/j.issn.1671－9638.20263102

·论著·

基于门诊病历的流行性感冒知识图谱构建与应用研究

戴萍萍[1]，陈泽华[1,2]，聂洋波[2]，彭 振[2]，姚志强[2]，许林勇[1,3]

[1. 中南大学湘雅三医院 湖南省高等学校医学信息研究重点实验室，湖南 长沙 410013；2. 中南大学生命科学学院，湖南 长沙 410083；3. 中南大学湘雅口腔医学院（湘雅口腔医院）口腔健康研究湖南省重点实验室，湖南 长沙 410008]

[摘 要] **目的** 比较不同深度学习算法模型与大语言模型在流行性感冒（流感）实体自动抽取任务中的性能，并基于 RoBERTa-BiLSTM-CRF 模型构建流感领域知识图谱。**方法** 基于 4 421 条医院流感门诊病历记录进行实体抽取及模型构建，通过 6 种机器学习算法及 2 种大语言模型进行模型性能比较分析，筛选最优模型完成三元组抽取。采用卡方检验验证实体关系组相关性，最终基于 RoBERTa-BiLSTM-CRF 模型，结合药物数据库匹配、相似度算法计算及人工审核实现实体统一，并利用 Neo4j 图数据库构建流感知识图谱。**结果** RoBERTa-BiLSTM-CRF 模型在门诊病历实体抽取中性能最优，准确率 0.931，召回率 0.934，F1 值 0.932；成功挖掘流感相关的检查、症状、疾病、药物、方法实体类别，发现 690 对有效关系对；最终基于挖掘的流感实体关系数据集构建可视化知识图谱。**结论** 本研究验证了 RoBERTa-BiLSTM-CRF 模型在门诊病历实体抽取任务中的显著优势，利用该模型构建了基于门诊病历的流感知识图谱，为医疗实体抽取及相关领域的研究提供了方法学参考与数据基础。

[关 键 词] 流行性感冒；门诊病历；知识图谱；实体抽取

[中图分类号] R181.3+2

Construction and application of influenza knowledge graph based on outpatient medical records

DAI Pingping[1], *CHEN Zehua*[1,2], *NIE Yangbo*[2], *PENG Zhen*[2], *YAO Zhiqiang*[2], *XU Linyong*[1,3] (1. *Key Laboratory of Medical Information Research of College of Hunan Province, The Third Xiangya Hospital of Central South University, Changsha* 410013, *China*; 2. *School of Life Sciences, Central South University, Changsha* 410083, *China*; 3. *Hunan Key Laboratory of Oral Health Research, Xiangya Stomatological Hospital and Xiangya School of Stomatology, Central South University, Changsha* 410008, *China*)

[**Abstract**] **Objective** To compare the performance of different deep learning algorithm models and large language models in the automatic extraction task of influenza entities, and construct knowledge graph in the domain of influenza based on the RoBERTa-BiLSTM-CRF model. **Methods** Entity extraction and model construction were performed based on 4 421 outpatient medical records of influenza from hospital. Through comparison and analysis of the performance of 6 machine learning algorithms and 2 large language models, the optimal model was screened out for triplet extraction. The correlation of entity-relationship pairs was validated with chi-square test. Based on RoBERTa-BiLSTM-CRF model, and combined with medication database matching, similarity algorithm calculation, and manual review, entity alignment was finally achieved. Influenza knowledge graph was constructed using the Neo4j graph database. **Results** RoBERTa-BiLSTM-CRF model achieved the best performance in the outpatient

[收稿日期] 2025－10－14

[基金项目] 湖南省重点研发计划项目“突发公共卫生事件早期预测预警关键技术攻关”（2023SK2005）；中华医学会医学教育研究课题项目“人工智能交叉背景下生物统计学课程研究”（2023B356）

[作者简介] 戴萍萍（1986－），女（土家族），湖南省慈利县人，助理研究员，主要从事医药信息管理与预防医学相关研究。陈泽华为共同第一作者。

[通信作者] 许林勇 E-mail:xulinyong@csu.edu.cn

medical record entity extraction, with an accuracy of 0.931, recall of 0.934, and F1 score of 0.932. Entity categories of influenza-related examinations, symptoms, diseases, medications, and methods were successfully extracted, and 690 effective relationship pairs were found out. Finally, a visualized knowledge graph was constructed based on the extracted influenza entity-related dataset. **Conclusion** This study validates the significant advantages of the RoBERTa-BiLSTM-CRF model in entity extraction from outpatient medical records, constructs an influenza knowledge graph based on the model, and provides methodological reference and data foundation for research in medical entity extraction and related fields.

[**Key words**] influenza; outpatient medical record; knowledge graph; entity extraction

流行性感冒(流感)传播迅速,人群易在短期内感染[1],给公众健康和生命安全带来严峻挑战[2]。在全球公共卫生治理体系重构与“健康中国”战略纵深推进的双重驱动下,医疗信息化转型催生出流感防控新需求,整合电子病历数据与构建标准化医学知识图谱是提升防控效果的重要技术路径[3-5]。深度学习技术的演进为电子病历数据的处理提供了新路径,突破了传统方法处理此类数据时的瓶颈[6-8]。机器学习技术的引入极大拓展了知识图谱构建的数据源范围[9-12],而人工智能技术的发展推动了知识图谱在医疗领域的广泛应用[13-16]。目前,电子病历数据仍以非结构化形式为主,开发利用电子病历数据,需要提取非结构化数据中的关键信息即实体,使之成为结构化数据进行存储利用,因此实体抽取是实现关键信息提取的重要步骤。尽管在智能化实体抽取与识别[17-19]、实体对齐[20-22],以及基于大语言模型的实体识别[23-28]方面已有不少研究,但针对流感知识图谱构建的研究仍较为缺乏。本文基于流感门诊病历进行实体抽取与知识图谱构建,旨在丰富医学知识图谱,构建应用场景和循证依据,促进病历知识整合,推动医学数据融合与挖掘理论发展,拓展人工智能在医学领域的应用,辅助医疗科学决策,同时助力公众健康教育。

1　对象与方法

1.1　*研究对象*　调取某三甲医院 2020—2023 年中文门诊电子病历,共涵盖 4 421 条数据。门诊电子病历中存在大量文本,实体抽取工作训练难度高、训练样本需求量大[29]。为保障实体抽取环节的质量,本研究用代码对文本数据开展切割处理工作,将文本转化为长度在 40 个字符以内的短文本。李会芝等[30]研究表明,患者基本信息、疾病诊断、基础病史、治疗信息、检查检验结果等字段,对流感研究与防控意义重大。因此,本文对数据集进行进一步筛选,剔除 MED_HISTORY(既往史)、BODY_EXAM(身体检查)、REC_SUGGEST(诊疗意见)、ANAMNESIS(主诉)、ILLNESS_DESC(现病史)这五个关键字段均为空白的数据,最终保留 2 642 条数据,并转化为短文本数据 27 403 条用于后续的实体自动抽取及知识图谱构建,其中采用简单随机抽样抽取 524 条病历数据用于模型对比试验。

1.2　*研究方法*

1.2.1　*数据标注*　将筛选后的数据按 8∶2 的比例划分为训练集和测试集,进行切割预处理后通过 label-studio 进行标注,共标注五类实体,分别是症状、疾病、检查、药物、治疗方法。

1.2.2　*模型评估方法*　实体抽取模型常见评估方法包括准确率、召回率与 F1 值,作为训练模型的性能参数,上述指标常用于评估模型的抽取结果和实际情况的差别。准确率是模型预测为正类的实体中实际正确的比例,准确率越高,模型预测出的实体可靠性越强;召回率是实际存在的实体中被模型正确预测出来的比例,召回率越高,模型遗漏实体越少;F1 值是准确率和召回率的调和平均数,综合反映两者表现,F1 值越接近 1,模型性能越好。计算公式如下:

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

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

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

其中,P 表示准确率,R 表示召回率。

参与对比的机器学习模型包括以下六种:BERT-CRF、RoBERTa-CRF、BERT-Softmax、RoBERTa-Softmax、BERT-BiLSTM-CRF、RoBERTa-BiLSTM-CRF,以及两种主流大语言模型:DeepSeek-v3、Qwen3-Max。为评估大语言模型在流感门诊病历实体抽取任务中的性能,本研究选取 DeepSeek-V3 与 Qwen3-Max 两款主流模型进行零

样本(zero-shot)测试。所有模型调用均采用官方应用程序编程接口(application programming interface,API),参数统一设置为:temperature = 0,top_p = 1,最大输出长度 512 词元(tokens)。提示词为:“你是一个专业的医学信息抽取系统。请从以下门诊病历中抽取出所有医学实体,并严格分类。实体类型:SYMPTOM(症状)、DISEASE(疾病)、EXAMINATION(检查)、MEDICATION(药物)、TREATMENT(治疗方法)。输出格式为 JSON,每个实体包含实体文本和类型。”对于无法解析的样本,记为未识别。所有结果经人工复核后计算准确率、召回率与 F1 值。通过对比各模型在电子病历实体抽取任务中的准确率、召回率以及 F1 值,探究其实际效能。

1.2.3 卡方检验 采用 Pearson 卡方检验对实体抽取完成后的结构化病历数据进行深度分析,按 $\alpha = 0.05$ 水平进行检验。若部分数据理论频数低于最小值 5,则不再对其进行后续检验。

1.3 知识图谱构建

1.3.1 实体抽取 通过评估对五大类实体的抽取效果验证 RoBERTa-BiLSTM-CRF 模型在实际应用中的有效性。

1.3.2 实体对齐 针对不同类型的实体,采用不同的实体对齐流程以高效规范诊疗数据。在药物类实体对齐方面,选用国家药品监督管理局药品数据库作为药物实体对齐的参考库,并采取专家调查法,完善实体对齐结果。在非药物类实体对齐方面,采用软件工程法、文献调研法和专家调查法统一相关实体。以《流行性感冒诊疗方案》文件为基准,建立症状术语与方法术语的标准化映射体系。通过机器筛选、人工标注和专家审核,保证标准化映射体系的规范性。运用 Python 代码中的 fuzzywuzzy 库,对抽取出的实体进行相似度计算,再引入人工审核环节,由专业人员与专家对初步对齐结果进行细致查验。

1.3.3 流感知识图谱的生成 将抽取的三元组与结构化数据生成的三元组进行对齐与冗余清洗后,利用 Neo4j 图数据库构建动态可扩展的存储模型,实现流感电子病历知识图谱的实体-关系可视化和语义化检索,完成知识图谱构建。

2 结果

2.1 模型性能评估与比较 与 BERT-CRF 模型(F1 值为 0.918)相比,BiLSTM 对 BERT 模型的显著性能增益。同时,RoBERTa 基础模型直接与 CRF 结合,能达到与 BERT-BiLSTM-CRF 相当的性能水平,提示 RoBERTa 文本表征方面拥有比 BERT 模型更强的能力。对比 BERT-Softmax 和 BERT-CRF 可见,BERT-Softmax 的 F1 值为 0.904,而 BERT-CRF 的 F1 值提升至 0.918,表明相较传统的 Softmax,加入 CRF 层能够更好地利用句子中相邻标签之间的依赖关系,确定实体边界。相比之下,大语言模型在实体抽取任务中的 F1 值仅为 0.567 和 0.495,显著低于所有基于 BERT 的对比模型,表明其在处理专业性强、实体边界模糊的门诊病历文本时尚存在明显局限。见表 1。

表 1 各模型实验结果对比

Table 1 Comparison of experimental results from different models

模型种类	P	R	F1
BERT-BiLSTM-CRF	0.921	0.933	0.927
BERT-CRF	0.911	0.925	0.918
BERT-Softmax	0.902	0.907	0.904
RoBERTa-Softmax	0.912	0.926	0.919
RoBERTa-BiLSTM-CRF	0.931	0.934	0.932
RoBERTa-CRF	0.919	0.934	0.927
Qwen3-Max	0.518	0.627	0.567
DeepSeek-V3	0.393	0.667	0.495

注:P 表示准确率,R 表示召回率。

疾病实体识别方面,大语言模型 F1 值与 RoBERTa-BiLSTM-CRF 模型相当,而 R 值高于 RoBERTa-BiLSTM-CRF 模型;其他实体识别方面,大语言模型的各项指标均远低于 RoBERTa-BiLSTM-CRF 模型。RoBERTa-BiLSTM-CRF 模型与大语言模型的各实体抽取性能对比试验结果见表 2。

表 2　RoBERTa-BiLSTM-CRF 模型与大语言模型对各实体类型抽取性能比较

Table 2　Comparison of the extraction performance of RoBERTa-BiLSTM-CRF model and large language models for various entity types

实体类型	RoBERTa-BiLSTM-CRF			DeepSeek-V3			Qwen3-Max		
	P	R	F1	P	R	F1	P	R	F1
症状	0.882	0.862	0.872	0.327	0.816	0.467	0.542	0.801	0.647
疾病	0.934	0.914	0.924	0.826	0.954	0.885	0.862	0.954	0.906
检查	0.921	0.901	0.911	0.521	0.521	0.521	0.496	0.435	0.464
药物	0.909	0.889	0.899	0.705	0.711	0.708	0.744	0.711	0.727
方法	0.867	0.847	0.857	0.150	0.422	0.221	0.265	0.409	0.322
总体	0.931	0.934	0.932	0.393	0.667	0.495	0.518	0.627	0.567

注：P 表示准确率，R 表示召回率。

2.2　基于 RoBERTa-BiLSTM-CRF 模型的实体自动抽取结果　对实体类型整体的 P 值为 0.950，R 值为 0.978，反映出对不同实体类型的处理整体处于较高水平，具备较好的性能和可靠性。大部分类别（除“方法”外）的 P 值和 R 值都较接近且均处于较高区间，提示对常见实体类型的处理能力相对均衡。F1 值方面，疾病（0.972）和检查（0.962）实体的 F1 值均超过 0.96，反映模型对疾病名称、检验检查项目等临床核心诊断要素具有较强的抽取能力。症状（0.941）与药物（0.929）实体的 F1 值处于中高水平，其性能可能受症状表述多样性及药物名称多源性的影响而出现一定波动。方法实体的 F1 值为 0.870，受医疗操作术语复杂性及标注细粒度差异的限制，性能相对偏低。见表 3。

表 3　模型训练结果

Table 3　Model training results

实体类型	P	R	F1
疾病	0.964	0.980	0.972
症状	0.931	0.951	0.941
药物	0.912	0.947	0.929
检查	0.954	0.969	0.962
方法	0.820	0.926	0.870
总体	0.950	0.978	0.964

注：P 表示准确率，R 表示召回率。

基于 RoBERTa-BiLSTM-CRF 模型进行实体自动抽取工作，并与人工标注的数据比对，剔除不合理的抽取结果以保证实体抽取工作的科学性和严谨性。抽取结果显示，电子病历内部数量最多的实体为检查（17 400 个），最少的为疾病（2 776 个）。见表 4。

表 4　各实体类型数量

Table 4　Number of each entity type

实体类型	数量
症状	15 770
疾病	2 776
检查	17 400
药物	4 623
方法	3 131

引入人工审核环节，由 4 名专业人员对初步对齐结果进行细致查验。同时，邀请相关领域专家给予针对性指导，依据专业知识与经验，对实体统一结果开展深度优化，实体对齐后，各类实体数量均显著减少。其中，检查类实体从 220 个减至 92 个，症状类从 814 个减至 246 个，疾病类从 52 个减至 45 个，药物类从 277 个减至 190 个，方法类从 114 个减至 57 个。

2.3　卡方检验结果　为系统挖掘流感门诊病历中各类实体之间的潜在关联，采用卡方检验对症状、药物、检查、治疗方法等实体类型关联频次之间的共现关系进行统计学检验，共筛选出具有统计学意义（$P \leqslant 0.05$）的潜在关联对 690 对。各类实体之间的关联频次见表 5。

表 5 实体关系关联频次表

Table 5 Association frequency of entity-relationship

实体类型 1	实体类型 2	关联频次
症状	方法	73
药物	检查	169
药物	方法	36
症状	检查	304
症状	药物	108

症状涵盖乏力、高热、咳嗽等，治疗方法包括口服药物治疗、密切观察、隔离等。大部分症状与密切观察关联（均 $P<0.001$），部分症状如偶咳等与口服药物治疗关联（$P=0.048$）。见表 6。

表 6 症状 - 方法卡方检验结果（部分）

Table 6 Symptom-method chi-square test results (Partial)

症状	方法	χ^2	P
乏力	密切观察	15.86	<0.001
高热	密切观察	19.73	<0.001
咽部红肿	及时复诊	7.88	0.005
咽部红肿	密切观察	14.27	<0.001
咽部红肿	隔离	12.39	<0.001
咽痛	密切观察	16.51	<0.001
咽痛	注意观察	4.45	0.035
高热	注意观察	16.16	<0.001
咳嗽	及时复诊	13.36	<0.001
偶咳	口服药物治疗	3.93	0.048

部分药物与特定检查项目关联，如头孢克肟类药物与单核细胞百分比、阿莫西林类药物与单核细胞百分比、蓝芩类药物与白细胞计数（均 $P\leqslant 0.05$），见表 7。

如表 8 所示，不同症状与特定药物存在关联，如乏力与头孢类药物，高热与小儿氨酚黄那敏类药物，咳嗽与奥司他韦类、小儿豉翘类等药物，头痛与头孢类药物的关联均具有统计学意义（均 $P<0.05$）。

表 7 药物 - 检查项目卡方检验结果（部分）

Table 7 Medication-examination chi-square test results (Partial)

药物	检查项目	χ^2	P
头孢克肟类药物	单核细胞百分比	3.84	0.050
阿莫西林类药物	单核细胞百分比	3.93	0.048
蓝芩类药物	中性粒细胞百分比	25.62	<0.001
蓝芩类药物	C-反应蛋白测定	28.95	<0.001
蓝芩类药物	白细胞计数	31.27	<0.001
蓝芩类药物	血小板计数	24.36	<0.001
小柴胡类药物	心率	19.71	<0.001
奥司他韦类药物	呼吸道合胞病毒	4.10	0.043
奥司他韦类药物	嗜酸细胞百分比	4.76	0.029

表 8 症状 - 药物卡方检验结果（部分）

Table 8 Symptom-medication chi-square test results (Partial)

症状	药物	χ^2	P
乏力	头孢类药物	7.24	0.007
乏力	小柴胡类药物	6.31	0.012
高热	奥司他韦类药物	5.02	0.025
高热	小儿氨酚黄那敏类药物	29.19	<0.001
高热	蓝芩类药物	4.53	0.033
偶咳	阿莫西林类药物	4.70	0.030
头晕	小柴胡类药物	4.06	0.044
头晕	布洛芬类药物	7.03	0.008
头晕	阿莫西林类药物	5.41	0.020
头痛	头孢类药物	4.96	0.026
咳嗽	奥司他韦类	8.81	0.003
咳嗽	小儿豉翘类	5.11	0.024

2.4 *知识图谱可视化* 根据 RoBERTa-BiLSTM-CRF 模型训练出的结果，以 LOAD CSV 格式导入数据，利用 Neo4j 的可视化属性，完成知识图谱的可视化操作。图谱显示，“流感”作为核心节点，向外发散出患者节点。通过患者节点，可以获取流感常见症状（如高热、咳嗽、乏力等）、药物和检查（如血常规、病毒核酸检测等），以及可能出现的并发症等信息。见图 1。

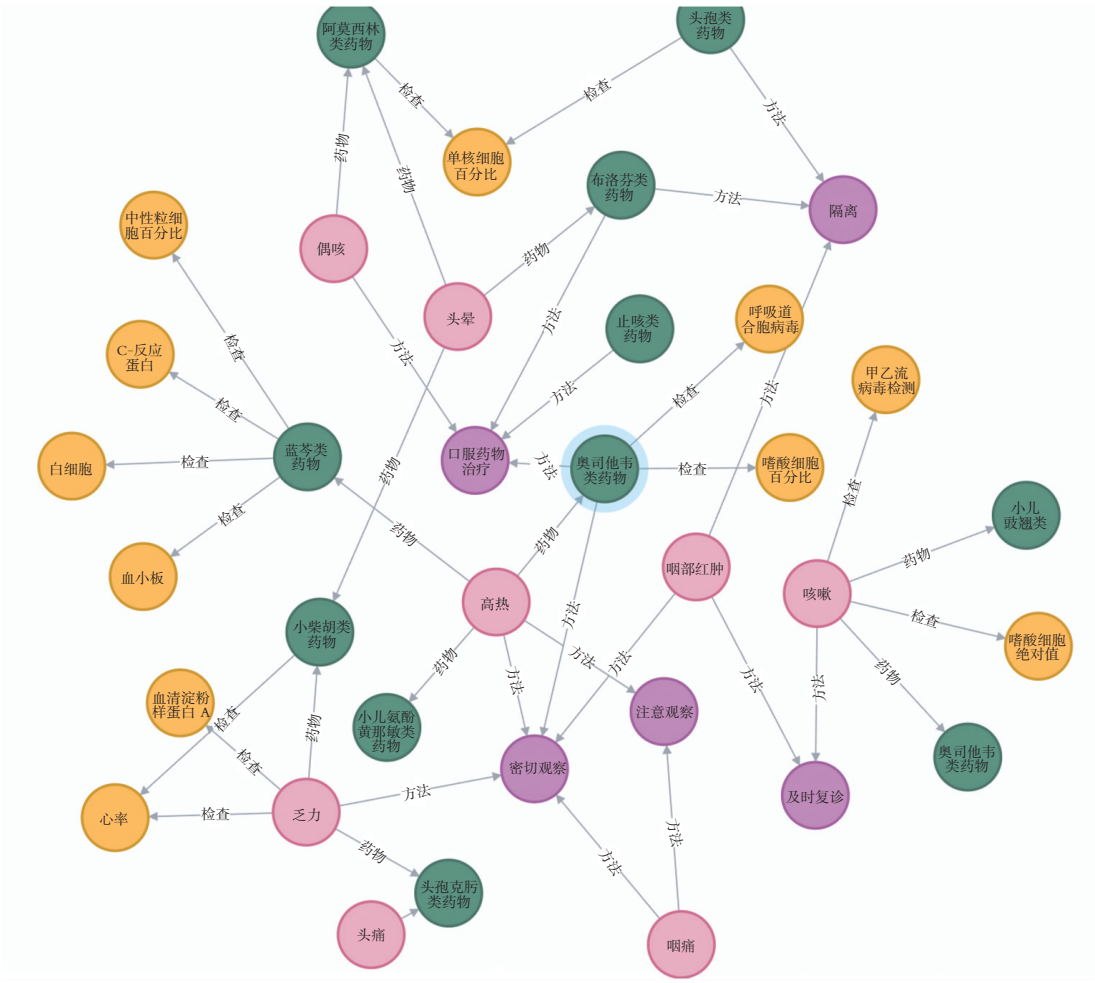


图 1　Neo4j 可视化效果图(仅展示部分实体关系)

Figure 1　Effectiveness of Neo4j visualization (Partial entity relationships)

3　讨论

3.1　*在流感门诊病历实体抽取任务中 RoBERTa-BiLSTM-CRF 模型性能最佳*　本研究基于真实门诊电子病历数据,聚焦流感实体抽取模型的比较分析,重点对比了 RoBERTa-BiLSTM-CRF 模型与其他主流模型以及大语言模型在架构、性能及应用于电子病历实体抽取场景时的差异。在性能上,该模型表现最佳,F1 值达 0.932,超越 BERT-BiLSTM-CRF、BERT-CRF 等多种对比模型,与研究[31-32]结果一致。结果表明,BiLSTM 对 BERT 模型性能提升显著,且 RoBERTa 在文本表征方面能力优于 BERT,CRF 层也能有效捕捉标签依赖关系,提高实体抽取准确性。尽管当前主流大语言模型在疾病实体识别上具有一定泛化能力[27-28],但在精确性、稳定性、计算效率等方面,基于 BERT 的模型仍具优势,表明实体抽取应用层面仍离不开基于 BERT 的模型。RoBERTa-BiLSTM-CRF 模型在流感门诊电子病历实体抽取任务中展现出的最佳性能,证实可基于此模型进行知识图谱的构建。杜文然等[33]同样验证了该模型在其他实体抽取中性能表现最佳。

3.2　*基于流感门诊病历的真实世界数据可抽取多种有效实体关系对*　本研究采用卡方检验,对系统挖掘出的症状、药物、检查、治疗方法等实体间的潜在关联进行验证,共发现 690 对有效关系,揭示了临床诊疗中各类实体间内在联系,为优化流感诊疗方案和合理用药提供了数据支持。

3.3　*流感知识图谱的构建可有效挖掘和展示不同实体间的关联*　通过构建流感知识图谱,将专业临床知识转化为通俗易懂且可视化的关联信息,可以帮助公众快速了解流感的症状识别、预防措施、合理用药等知识,提升全民流感防控意识,不仅系统整合了流感临床知识,还在临床诊疗、科研创新、公共卫

生防控及健康管理等多场景中体现实用价值，为流感诊疗智能化、防控精准化提供了有力支撑。

3.4 研究的局限性 数据仅来源于单一中心。由于电子病历的书写习惯、诊疗规范存在地域和机构差异，可能影响本研究在不同地域、医疗环境数据中的泛化能力。知识图谱中的实体关系本质上是基于真实世界门诊病历数据通过卡方检验挖掘的统计共现关联，并非确定的临床因果或治疗关系，需结合医学专业知识审慎解读。

未来研究将聚焦于多中心数据验证与融合，以及模型性能优化与功能拓展。一方面，拓展更具多样性和代表性的数据源，开展跨机构、跨地域的多中心合作。另一方面，拟引入图神经网络（graph neural network，GNN）改进现有 RoBERTa-BiLSTM-CRF 模型，积极探索多模态数据融合路径，将医学影像、语音诊疗记录等纳入模型训练体系，构建动态更新机制，不断完善与升级知识图谱。

综上所述，RoBERTa-BiLSTM-CRF 模型适用于流感门诊病历的实体抽取任务，病历数据的结构化处理，为知识图谱的构建奠定了坚实基础，有效丰富图谱内容，在流感防控、公众教育等方面展现出巨大应用潜力。后续将致力于深化与拓展系统在临床与公共卫生领域的应用，联合科研机构、药企等多方力量，协同推动流感相关研究，及时将前沿研究成果融入知识图谱，形成产学研用紧密结合、相互促进的良性发展格局。

利益冲突：所有作者均声明不存在利益冲突。

[参 考 文 献]

[1] 唐获音，丁奕州，王轩，等. 知识图谱技术在预测与健康管理中的应用现状与研究展望[J]. 电光与控制，2024，31(2)：1－11.
Tang DY，Ding YZ，Wang X，et al. Research progress and prospects of knowledge graph technology applied in prognostics and health management [J]. Electronics Optics & Control，2024，31(2)：1－11.

[2] 郑鑫彤，李燕，庞建美，等. 知识图谱在护理领域应用的研究进展[J]. 中国护理管理，2024，24(12)：1910－1913.
Zheng XT，Li Y，Pang JM，et al. Research progress on knowledge graph application in nursing field [J]. Chinese Nursing Management，2024，24(12)：1910－1913.

[3] 崔雷. 医学信息学专业教育中课程设计的理念[J]. 医学信息学杂志，2022，43(3)：15－17，39.
Cui L. The concept of curriculum design of education in medical informatics specialty[J]. Journal of Medical Informatics，2022，43(3)：15－17，39.

[4] Zhao H，Xiong WJ. A multi-scale embedding network for unified named entity recognition in Chinese Electronic Medical Records[J]. Alexandria Eng J，2024，107：665－674.

[5] 周育忠，林正平，王昕，等. 未来标准语义知识库的发展趋势[J]. 科技与创新，2025(1)：90－92.
Zhou YZ，Lin ZP，Wang X，et al. Development trend of future standard semantic knowledge base[J]. Science and Technology & Innovation，2025(1)：90－92.

[6] Miran SM，Nelson SJ，Redd D，et al. Using multivariate long short-term memory neural network to detect aberrant signals in health data for quality assurance[J]. Int J Med Inform，2021，147：104368.

[7] 余本功，张连彬. 基于 CP-CNN 的中文短文本分类研究[J]. 计算机应用研究，2018，35(4)：1001－1004.
Yu BG，Zhang LB. Research on Chinese short text classification based on CP-CNN[J]. Application Research of Computers，2018，35(4)：1001－1004.

[8] 姚琼，王觅也，师庆科，等. 深度学习在现代医疗领域中的应用[J]. 计算机系统应用，2022，31(4)：33－46.
Yao Q，Wang MY，Shi QK，et al. Application of deep learning in modern healthcare[J]. Computer Systems & Applications，2022，31(4)：33－46.

[9] 邹玉薇. 中文医疗知识图谱半自动化构建研究[D]. 武汉：武汉科技大学，2016.
Zou YW. Research on semi-automatic construction of Chinese medical knowledge graph[D]. Wuhan：Wuhan University of Science and Technology，2016.

[10] Zhu YJ，Che C，Jin B，et al. Knowledge-driven drug repurposing using a comprehensive drug knowledge graph [J]. Health Informatics J，2020，26(4)：2737－2750.

[11] 谭州科，白志勋. 知识图谱和人工智能在现代医学中的应用[J]. 教育教学论坛，2024(45)：21－24.
Tan ZK，Bai ZX. The application of knowledge graphs and artificial intelligence in modern medicine[J]. Education and Teaching Forum，2024(45)：21－24.

[12] 李宏伟，严武. 基于临床分布式大数据知识图谱的设计与应用研究[J]. 中国医疗管理科学，2023，13(3)：36－41.
Li HW，Yan W. Design and application of clinically distributed big data knowledge map[J]. Chinese Journal of Medical Management Sciences，2023，13(3)：36－41.

[13] Gao S，Yu KX，Yang Y，et al. Large language model powered knowledge graph construction for mental health exploration [J]. Nat Commun，2025，16(1)：7526.

[14] 施成谕. 基于知识图谱的合理用药分析与推荐方法研究[D]. 北京：北京工业大学，2022.
Shi CY. Analysis and recommendation of rational drug use based on knowledge graph[D]. Beijing：Beijing University of Technology，2022.

[15] 张晗，安欣宇，刘春鹤. 基于多源语义知识图谱的药物知识发现：以药物重定位为实证[J]. 数据分析与知识发现，2022，6

(7)：87－98.
Zhang H，An XY，Liu CH. Building multi-source semantic knowledge graph for drug repositioning[J]. Data Analysis and Knowledge Discovery，2022，6(7)：87－98.

[16] 谭威. 基于知识图谱的智能医疗问答及导诊系统的研究[D]. 上海：上海交通大学，2022.
Tan W. Research on intelligent medical question-answering and guidance system based on knowledge graph[D]. Shanghai：Shanghai Jiao Tong University，2022.

[17] Chapman AB，Peterson KS，Alba PR，et al. Detecting adverse drug events with rapidly trained classification models[J]. Drug Saf，2019，42(1)：147－156.

[18] 谭平，刘惠娜，韦昌法. 基于 BiLSTM-CRF 和 Neo4j 的脾胃病知识图谱构建[J]. 科学技术与工程，2025，25(22)：9436－9444.
Tan P，Liu HN，Wei CF. Spleen and stomach disease knowledge graph construction based on BiLSTM-CRF and Neo4j[J]. Science Technology and Engineering，2025，25(22)：9436－9444.

[19] 吴欢，车贺宾，王万玲，等. 基于循证医学和电子病历数据的通用医学知识图谱构建[J]. 医学信息学杂志，2025，46(2)：22－28.
Wu H，Che HB，Wang WL，et al. Construction of a general medical knowledge graph based on evidence-based medicine and electronic medical record data[J]. Journal of Medical Informatics，2025，46(2)：22－28.

[20] Qi ZY，Zhang ZH，Chen JY，et al. Unsupervised knowledge graph alignment by probabilistic reasoning and semantic embedding[C]//Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. California：IJCAI，2021：2019－2025.

[21] Liu X，Hong HY，Wang XH，et al. SelfKG：self-supervised entity alignment in knowledge graphs[C]//Proceedings of the ACM Web Conference 2022. New York，NY，USA：ACM，2022：860－870.

[22] 丁瑞卿，赵俊峰，王乐业. 综合实体语义和本体信息的多源中文医疗知识图谱实体对齐[J]. 软件学报，2025，36(11)：5178－5196.
Ding RQ，Zhao JF，Wang LY. Multi-source Chinese medical knowledge graph entity alignment via entity semantics and ontology information[J]. Journal of Software，2025，36(11)：5178－5196.

[23] Zhou JY，Ma ZL. Named entity recognition for construction documents based on fine-tuning of large language models with low-quality datasets[J]. Autom Constr，2025，174：106151.

[24] Zhang YL，Liu JJ，Zhong XF，et al. SecLMNER：a framework for enhanced named entity recognition in multi-source cybersecurity data using large language models[J]. Expert Syst Appl，2025，271：126651.

[25] Wang ZH，Chen HR，Xu G，et al. A novel large-language-model-driven framework for named entity recognition[J]. Inf Process Manag，2025，62(3)：104054.

[26] Lu J，Li F，Feng L，et al. Microllm：a structured information extraction tool using large language models and named entity recognition in microbiology[J]. Brief Bioinform，2025，26(5)：bbaf534.

[27] Liu YX，Bian JY，Zhai WQ，et al. Vaner2：towards more general biomedical named entity recognition using multi-task large language model encoders[J]. Health Inf Sci Syst，2025，14(1)：4.

[28] 杨波，孙晓虎，党佳怡，等. 面向医疗问答系统的大语言模型命名实体识别方法[J]. 计算机科学与探索，2023，17(10)：2389－2402.
Yang B，Sun XH，Dang JY，et al. Named entity recognition method of large language model for medical question answering system[J]. Journal of Frontiers of Computer Science and Technology，2023，17(10)：2389－2402.

[29] 赵继贵，钱育蓉，王魁，等. 中文命名实体识别研究综述[J]. 计算机工程与应用，2024，60(1)：15－27.
Zhao JG，Qian YR，Wang K，et al. Survey of Chinese named entity recognition research[J]. Computer Engineering and Applications，2024，60(1)：15－27.

[30] 李会芝，侯彩娥，任正伟，等. 电子病历档案在突发公共卫生事件期间的价值研究[J]. 办公室业务，2024(23)：187－189.
Li HZ，Hou CE，Ren ZW，et al. Study on the value of electronic medical records during public health emergencies[J]. Office Operations，2024(23)：187－189.

[31] 熊回香，周明洁. 电子病历中基于实体识别和共现分析的疾病间语义关系挖掘研究[J]. 情报科学，2025，43(6)：14－27.
Xiong HX，Zhou MJ. Semantic relationship mining between diseases based on entity recognition and co-occurrence analysis in Chinese electronic medical records[J]. Information Science，2025，43(6)：14－27.

[32] 袁贞明，沈辉，俞凯，等. 基于电子病历文本的诊疗事件实体抽取研究[J]. 中国数字医学，2021，16(7)：33－38.
Yuan ZM，Shen H，Yu K，et al. Research on the entity extraction of diagnosis and treatment events based on texts of electronic medical records[J]. China Digital Medicine，2021，16(7)：33－38.

[33] 杜文然，靳征，代明睿，等. 基于 RoBERTa-BiLSTM-CRF 模型的铁路货运一口价议价策略命名实体识别[J]. 铁路计算机应用，2023，32(5)：11－15.
Du WR，Jin Z，Dai MR，et al. Named entity recognition of railway freight competitive pricing strategy based on RoBERTa-BiLSTM-CRF model[J]. Railway Computer Application，2023，32(5)：11－15.

（本文编辑：翟若南）

本文引用格式：戴萍萍，陈泽华，聂洋波，等. 基于门诊病历的流行性感冒知识图谱构建与应用研究[J]. 中国感染控制杂志，2026，25(4)：566－573. DOI：10.12138/j.issn.1671－9638.20263102.

Cite this article as：DAI Pingping，CHEN Zehua，NIE Yangbo，et al. Construction and application of influenza knowledge graph based on outpatient medical records[J]. Chin J Infect Control，2026，25(4)：566－573. DOI：10.12138/j.issn.1671－9638.20263102.