



自然语言处理

在线峰会

产业创新与实践 论坛

2021.07.10 (周六) 09:00~17:30





融合领域知识的 医疗推理技术

苏嘉 博士 华为云语音语义创新Lab



目录

CONTENTS

01

智慧医疗现状

Subject

02

核心问题

Subject

03

领域知识

Subject

04

医疗知识计算

Subject

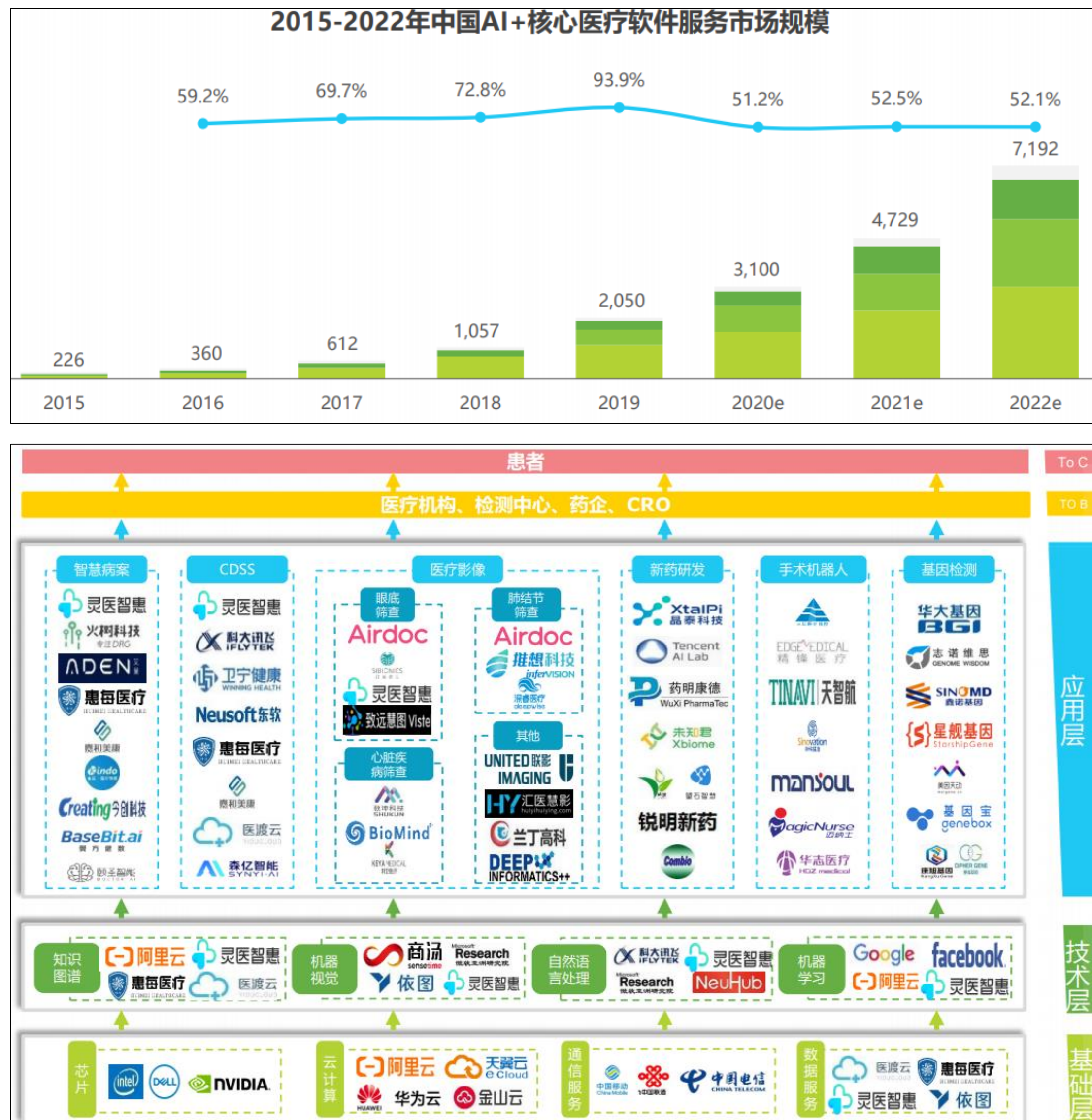
01 智慧医疗现状

Subject

前景
技术
重要事件

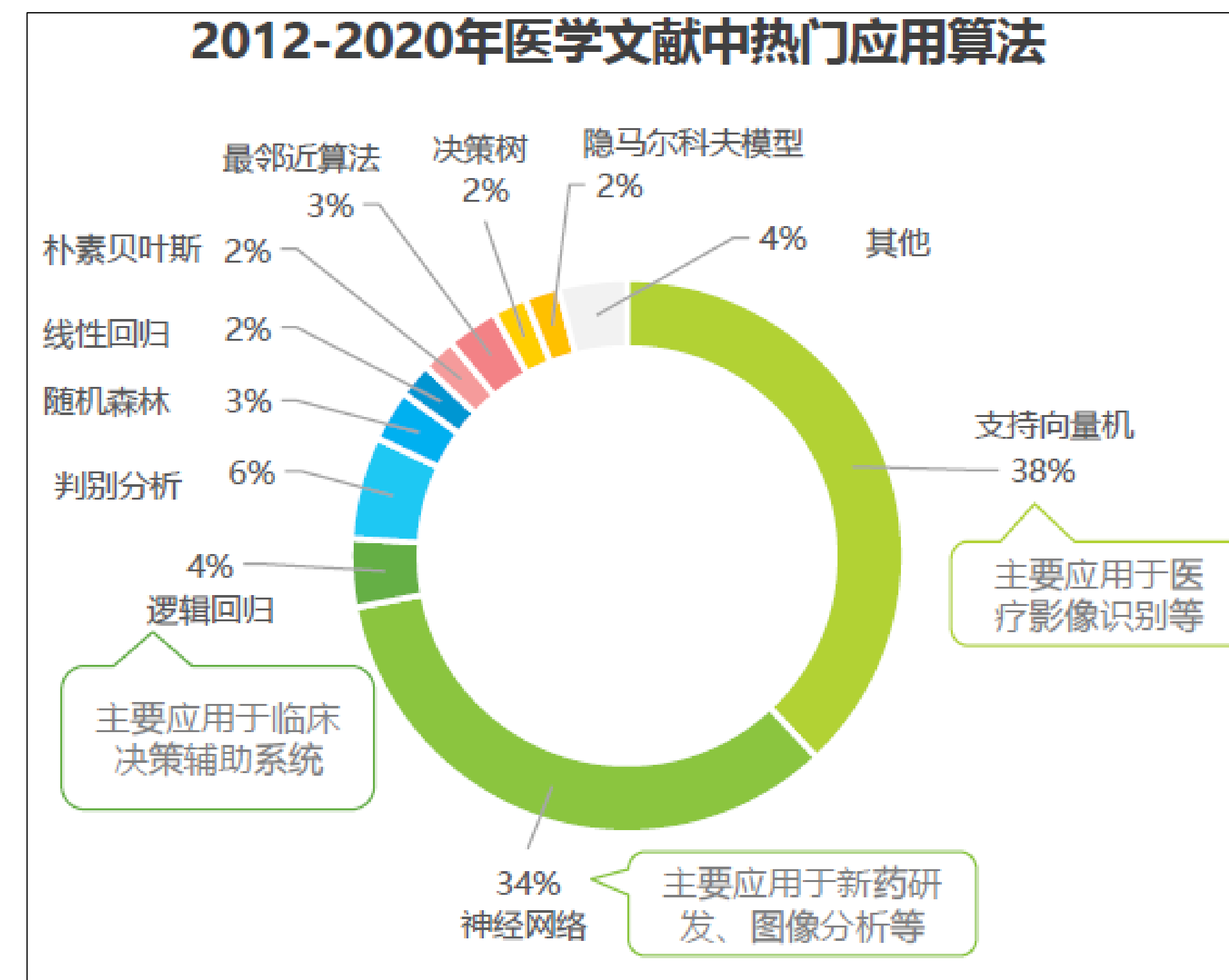
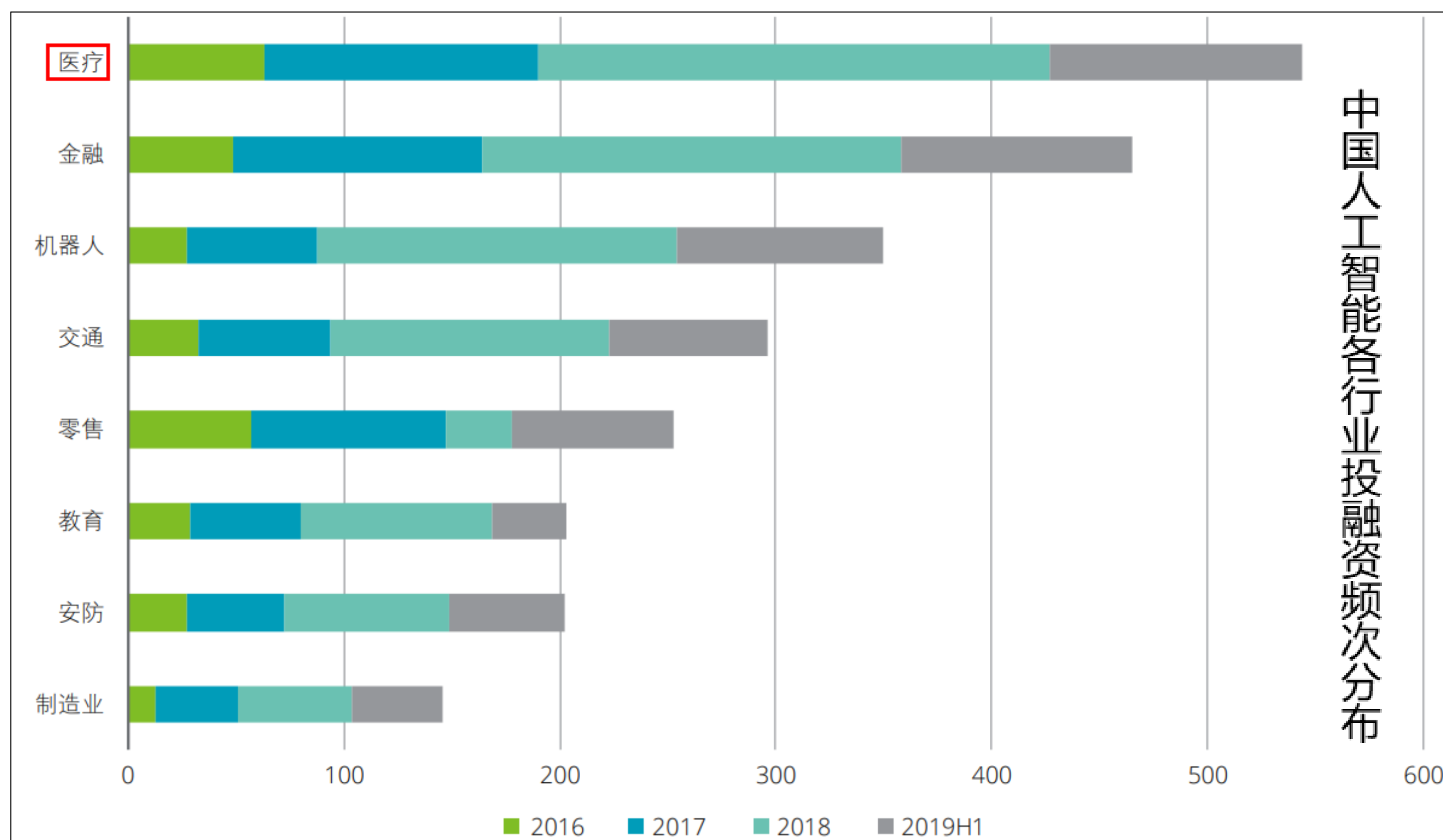
智慧医疗现状

- 市场规模：2025年全球276亿美元
- 全球发展格局：中美两国智慧医疗发展双足鼎力，日本、英国、以色列紧随其后
- 国内现状：
 - ✓ 医疗资源分布不均衡
 - ✓ 感知到认知智能的过渡
 - ✓ AI医学影像率先落地
 - ✓ 新冠疫情刺激智慧医疗行业发展
 - ✓ 政策支持
 - ✓ 智慧医疗产业



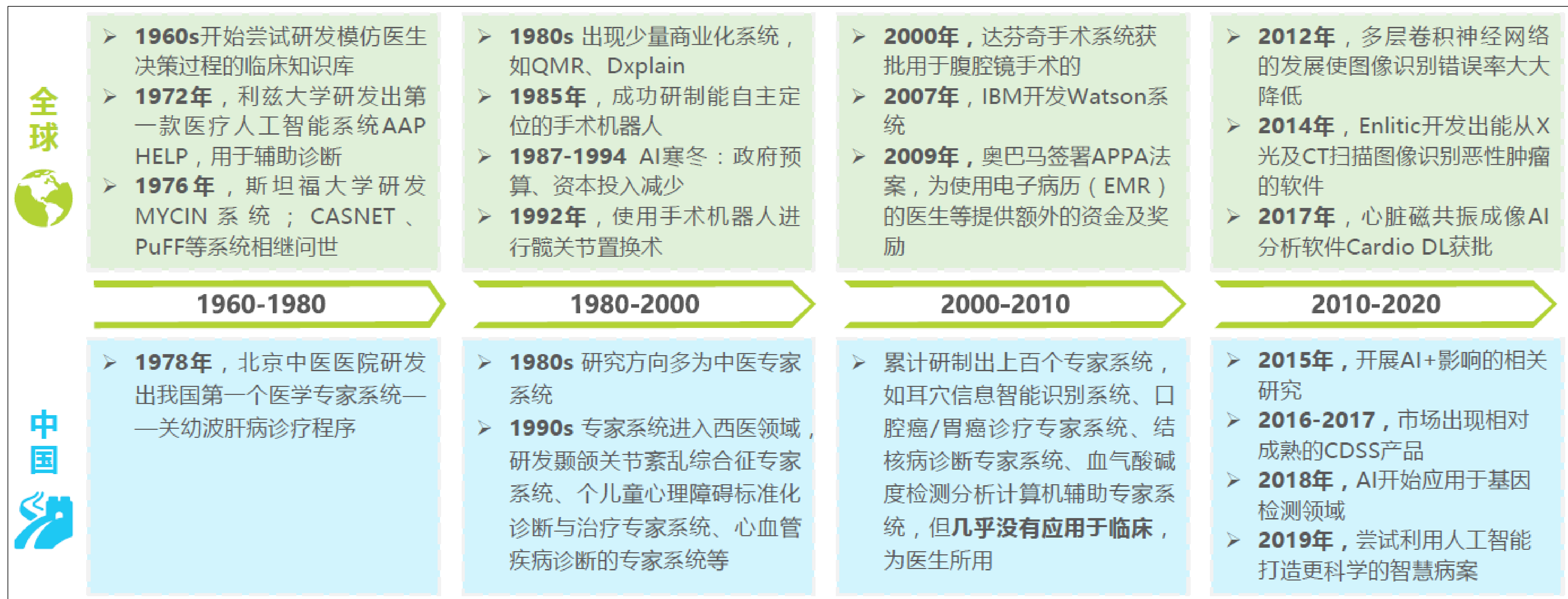
智慧医疗现状

较为成熟场景：疾病筛查和预测、医学影像诊断、临床**辅助**决策、病历与文献分析、新药发现
AI技术：计算机视觉、自然语言处理、语音识别、机器学习等



智慧医疗现状

重要事件时间线



02 核心问题

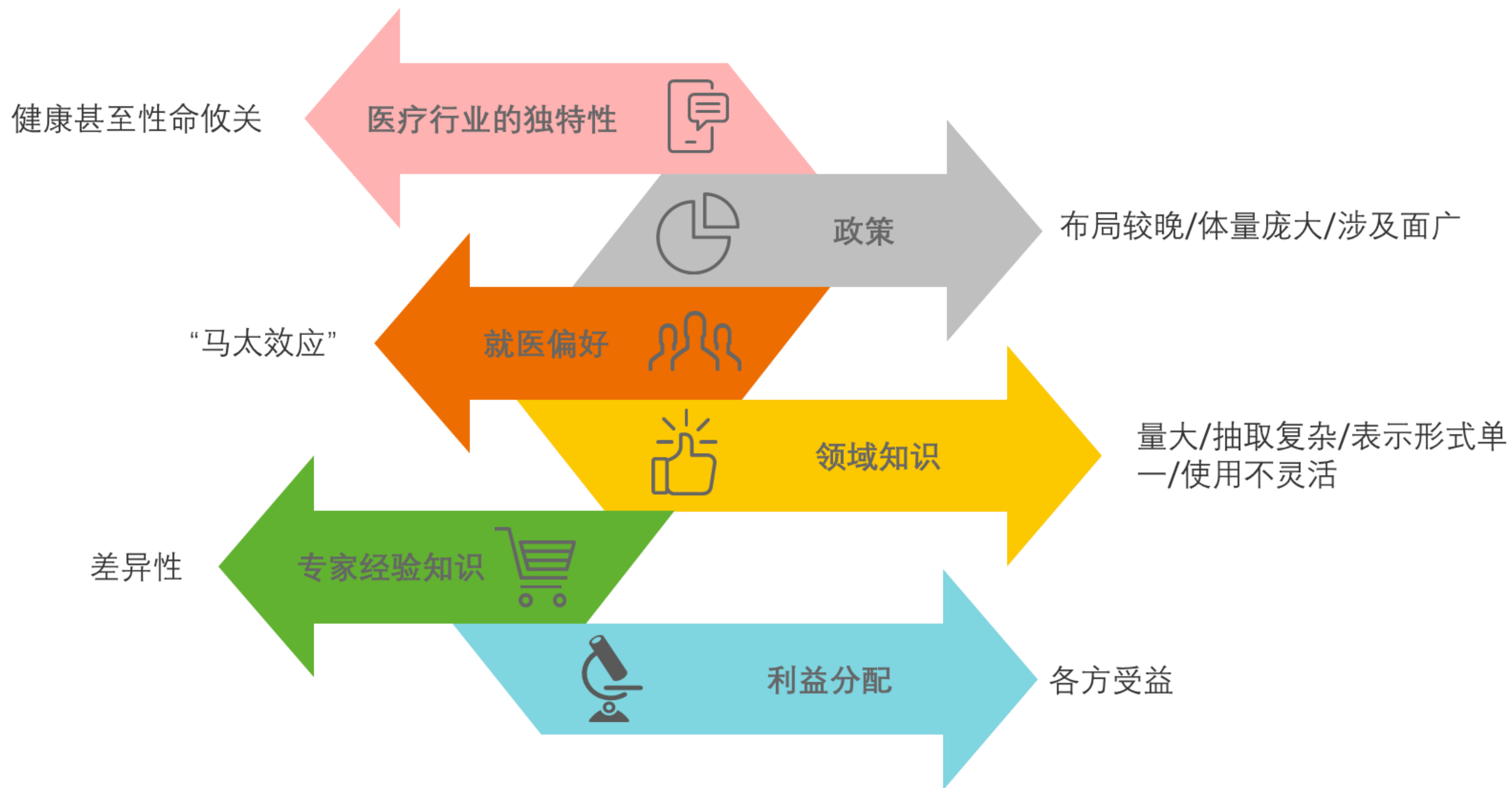
Subject



行业限制
数据
算法&算力

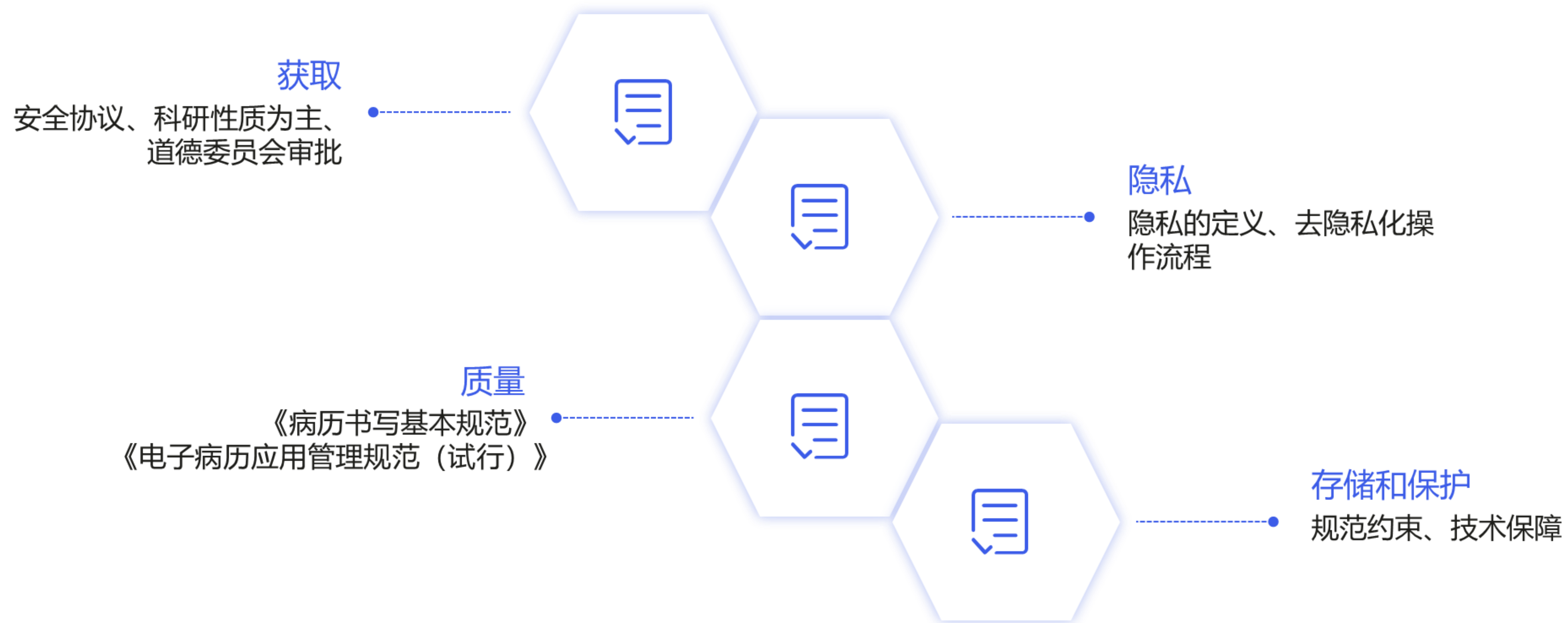
核心问题

行业限制



核心问题

数据



核心问题

算法&算力

可解释性：当前的AI模型近似黑盒

医生：希望获得推理证据链，有论证出处

患者：希望有出处和相似病例

个性化：

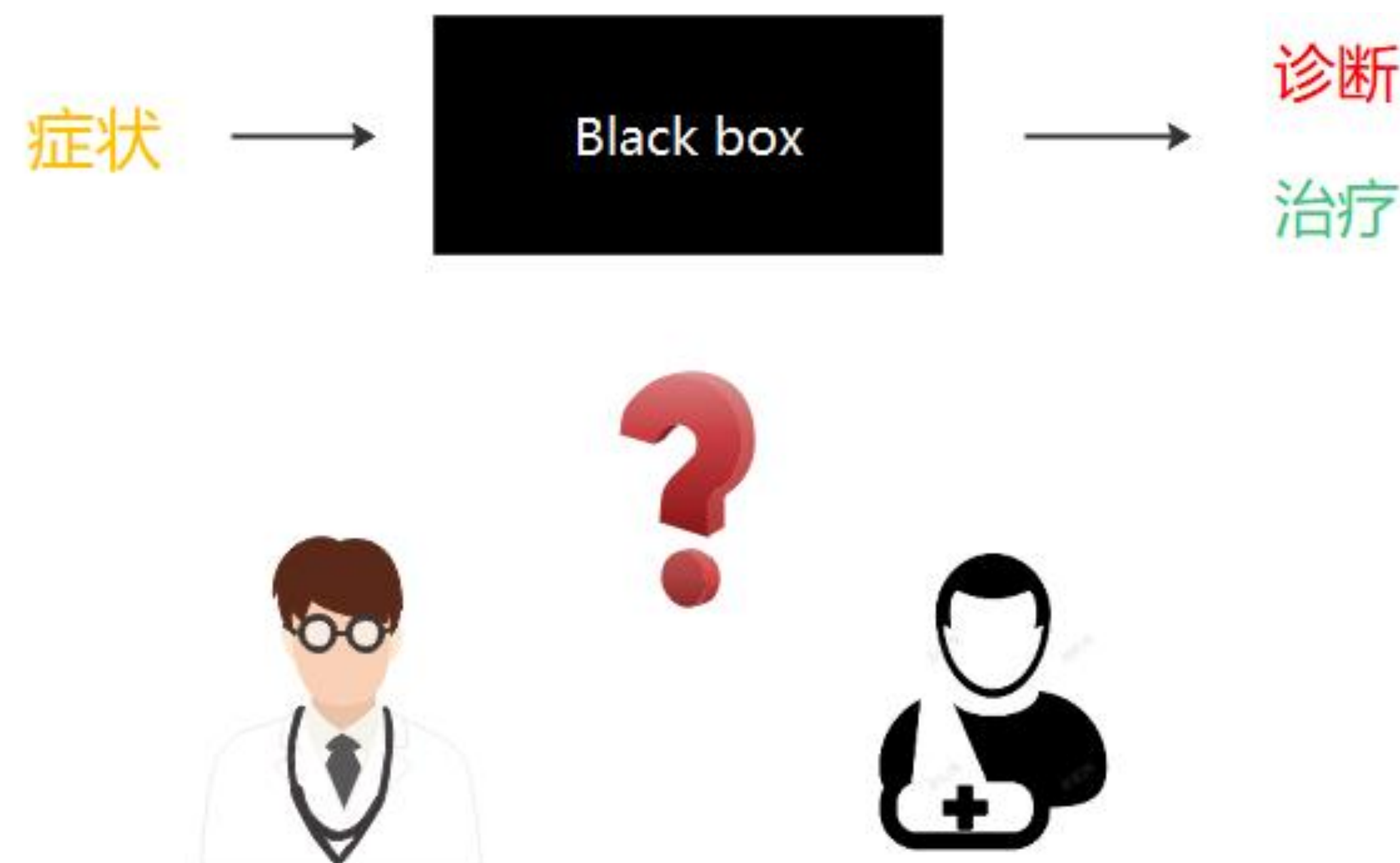
在数据不充沛的情况下，难以做到输入小的改动有差异化的输出

计算能力：

大模型的算力消耗

计算设备更新

从技术角度来看，如何将行业知识融入到算法当中
参与知识计算是核心



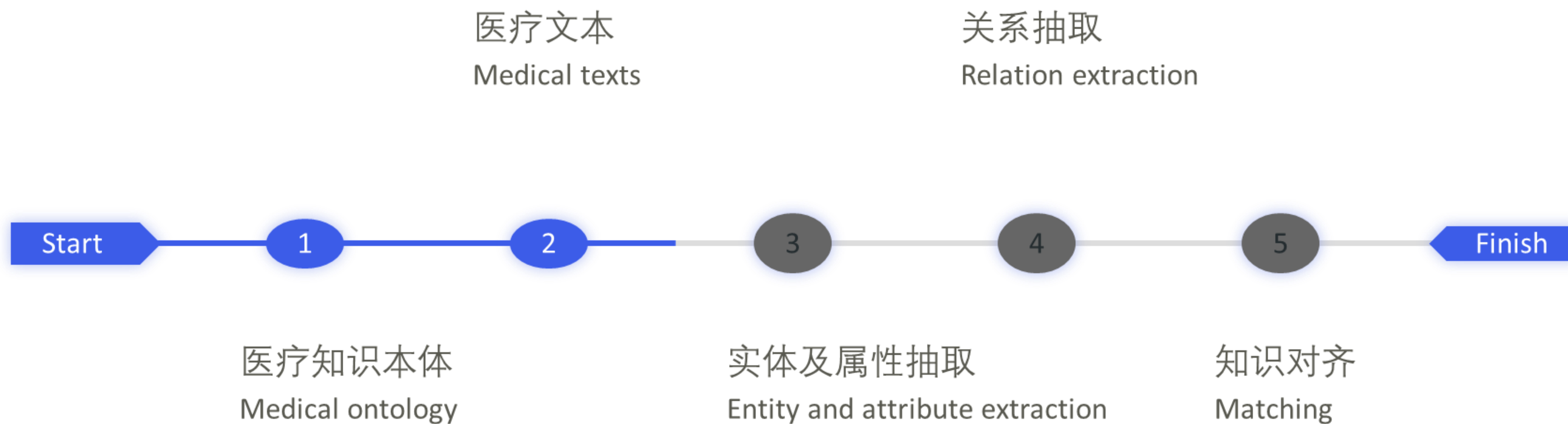
03 领域知识



医疗知识图谱，医疗预训练模型

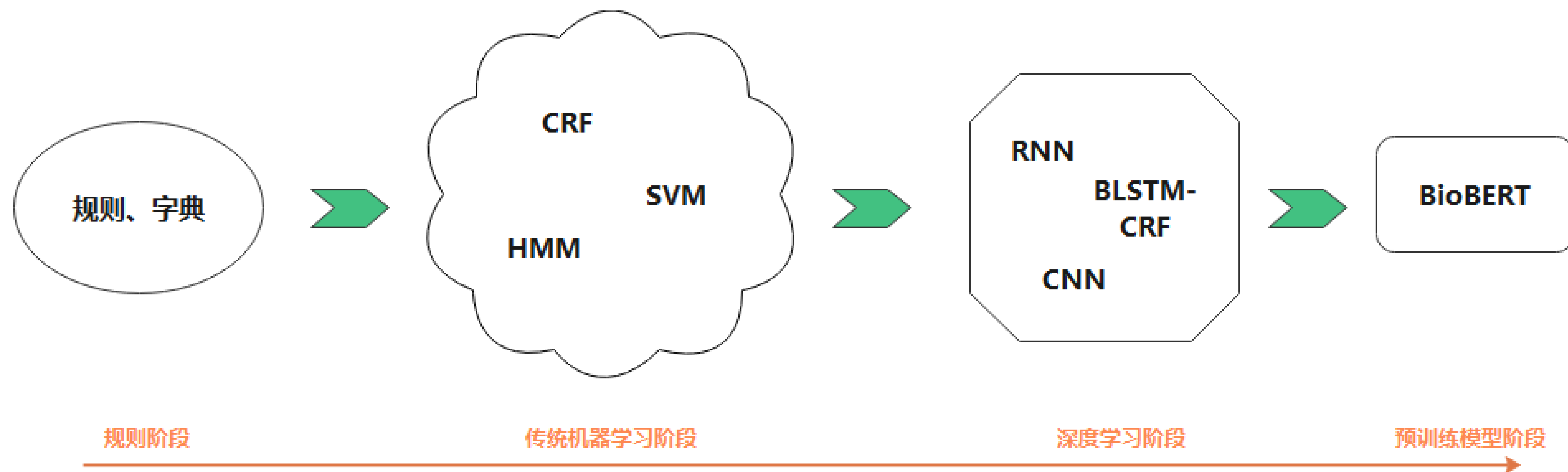
医疗知识图谱

构建



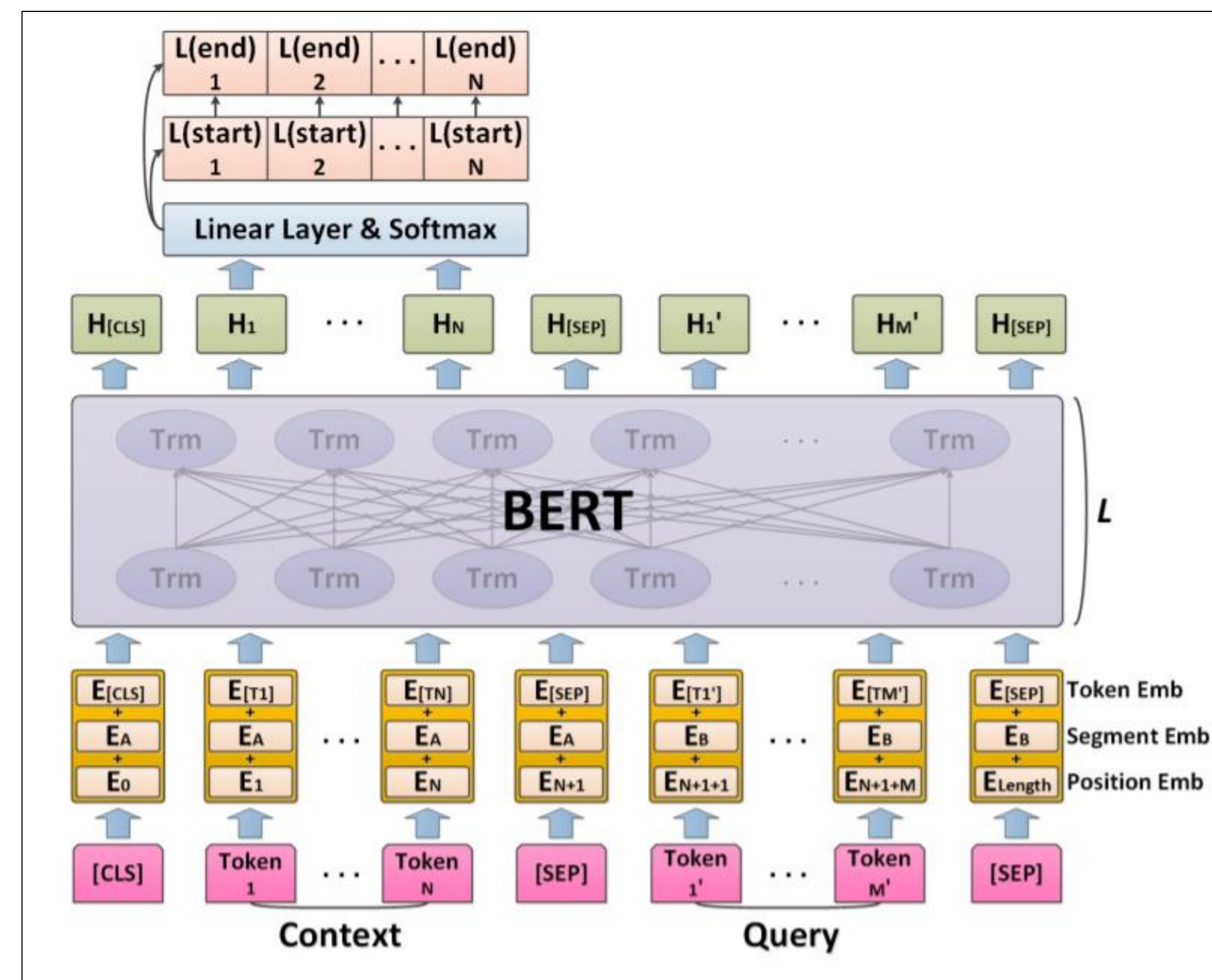
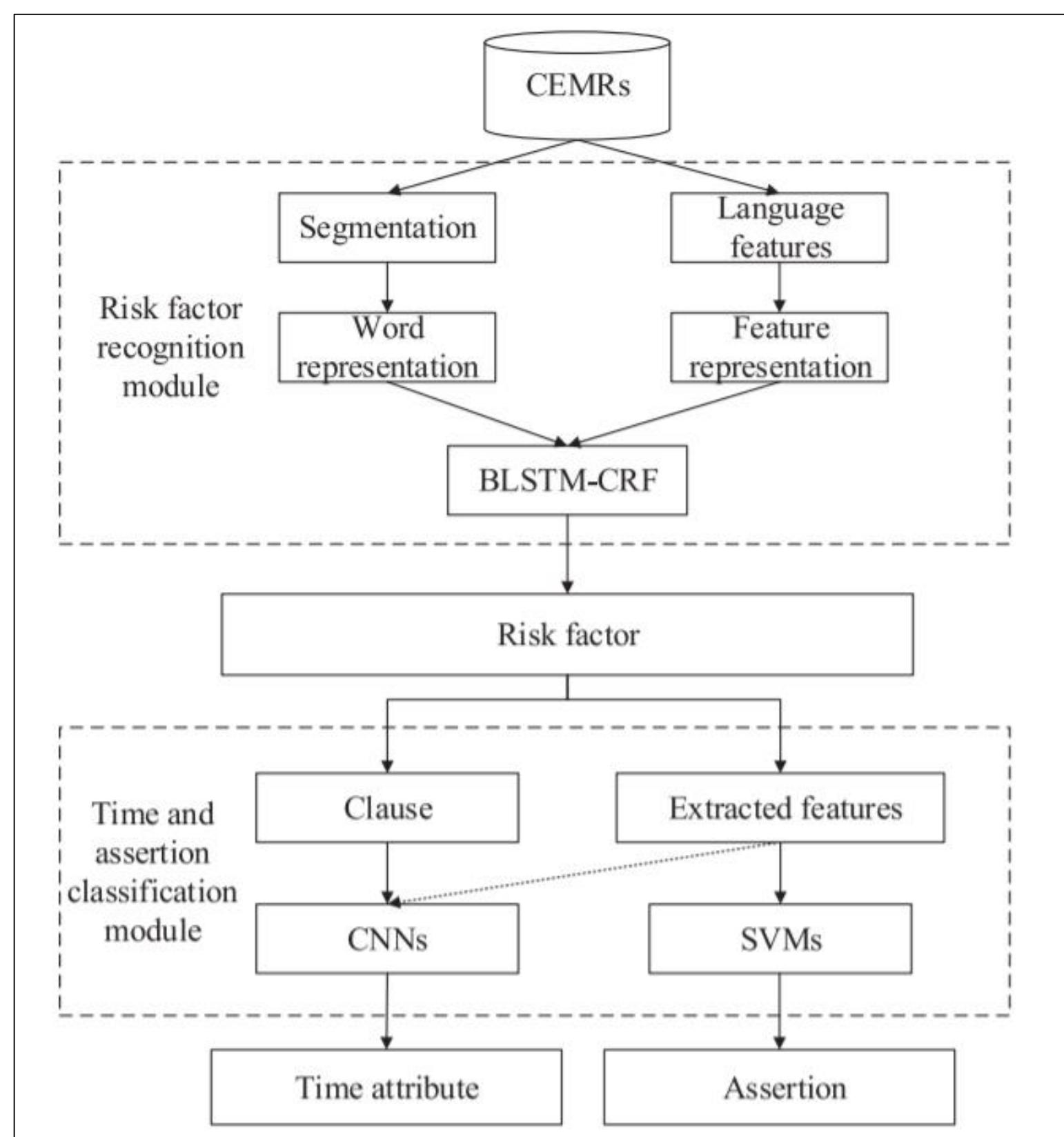
医疗知识图谱

抽取方法



医疗知识图谱

实体及属性识别

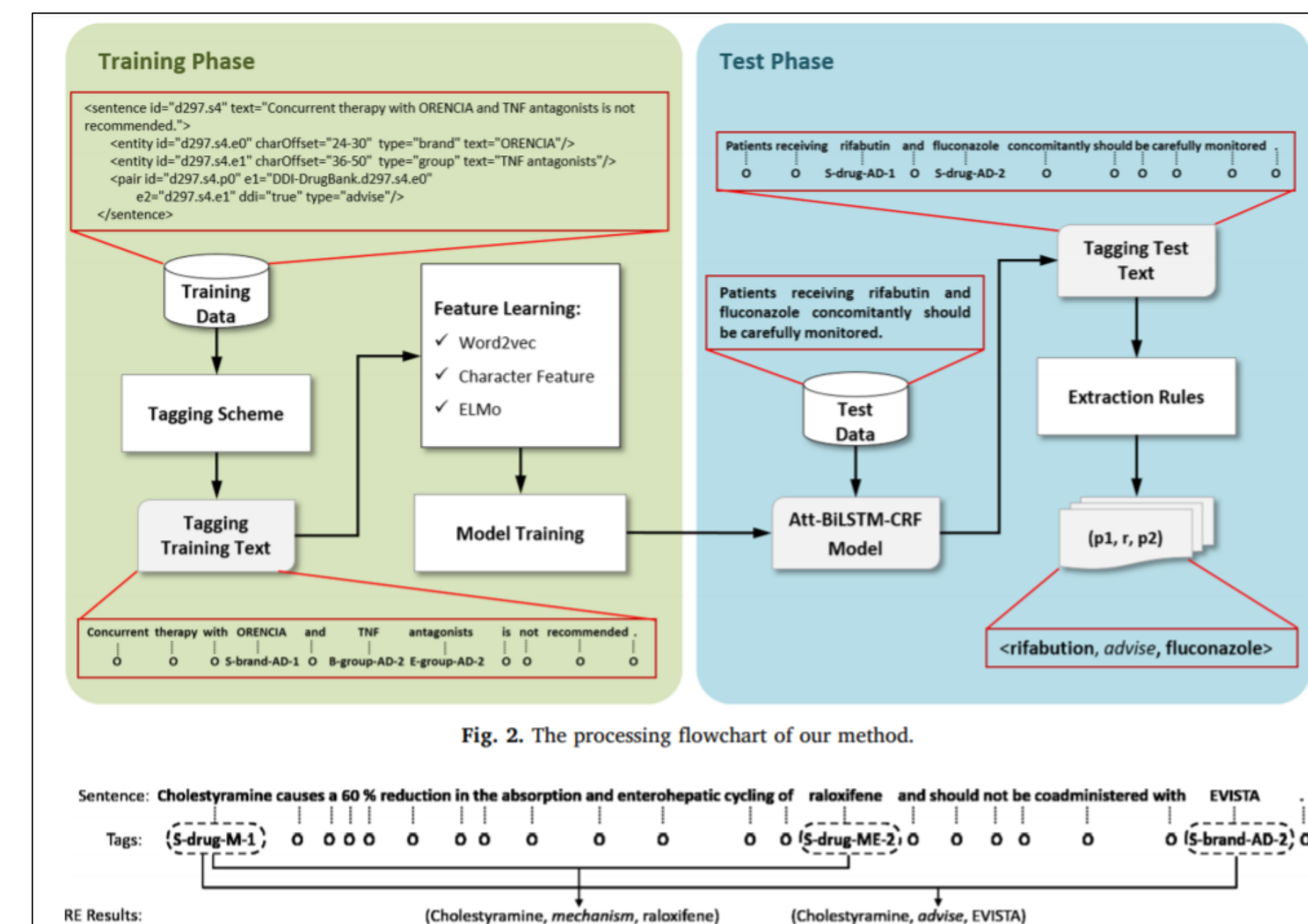
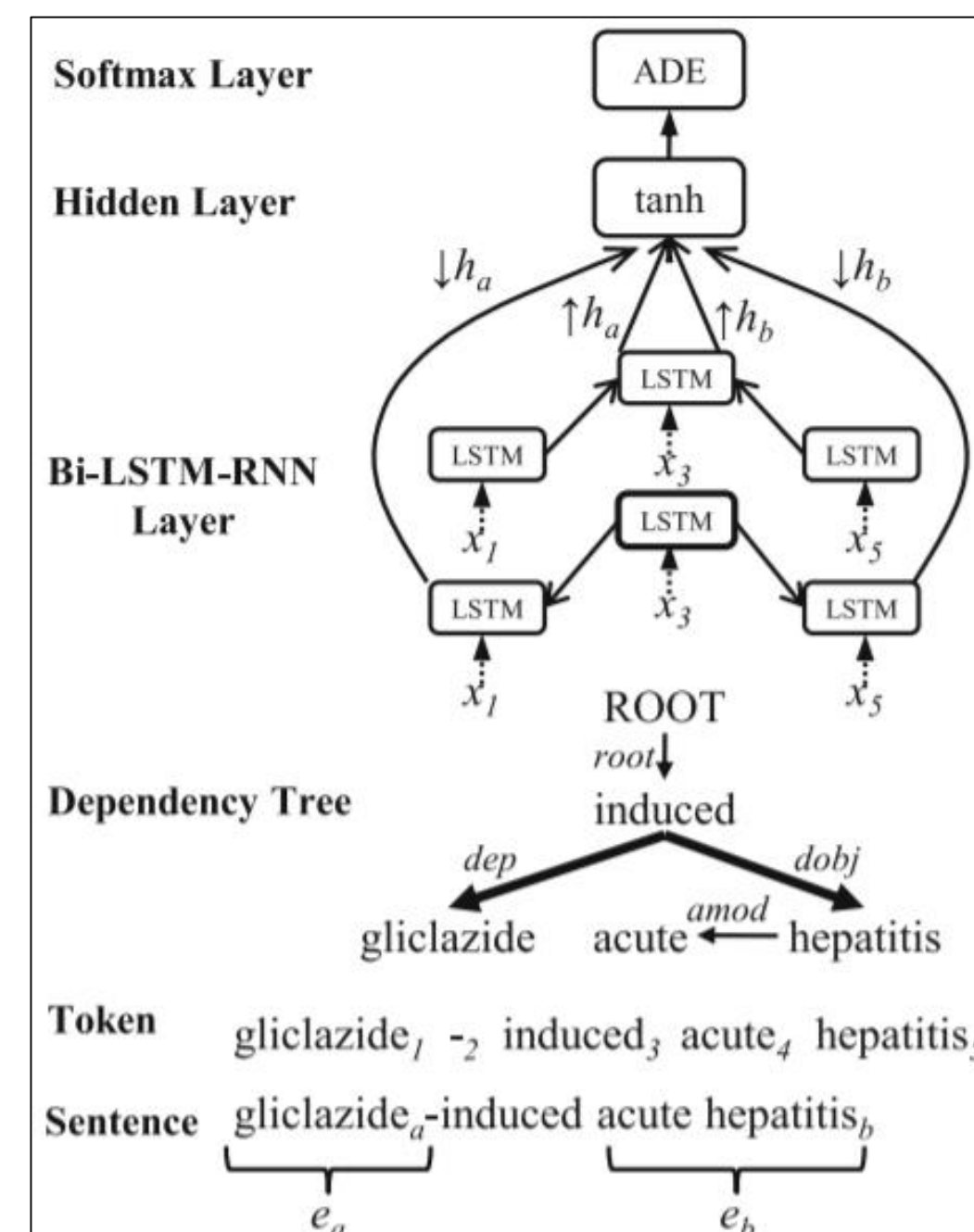
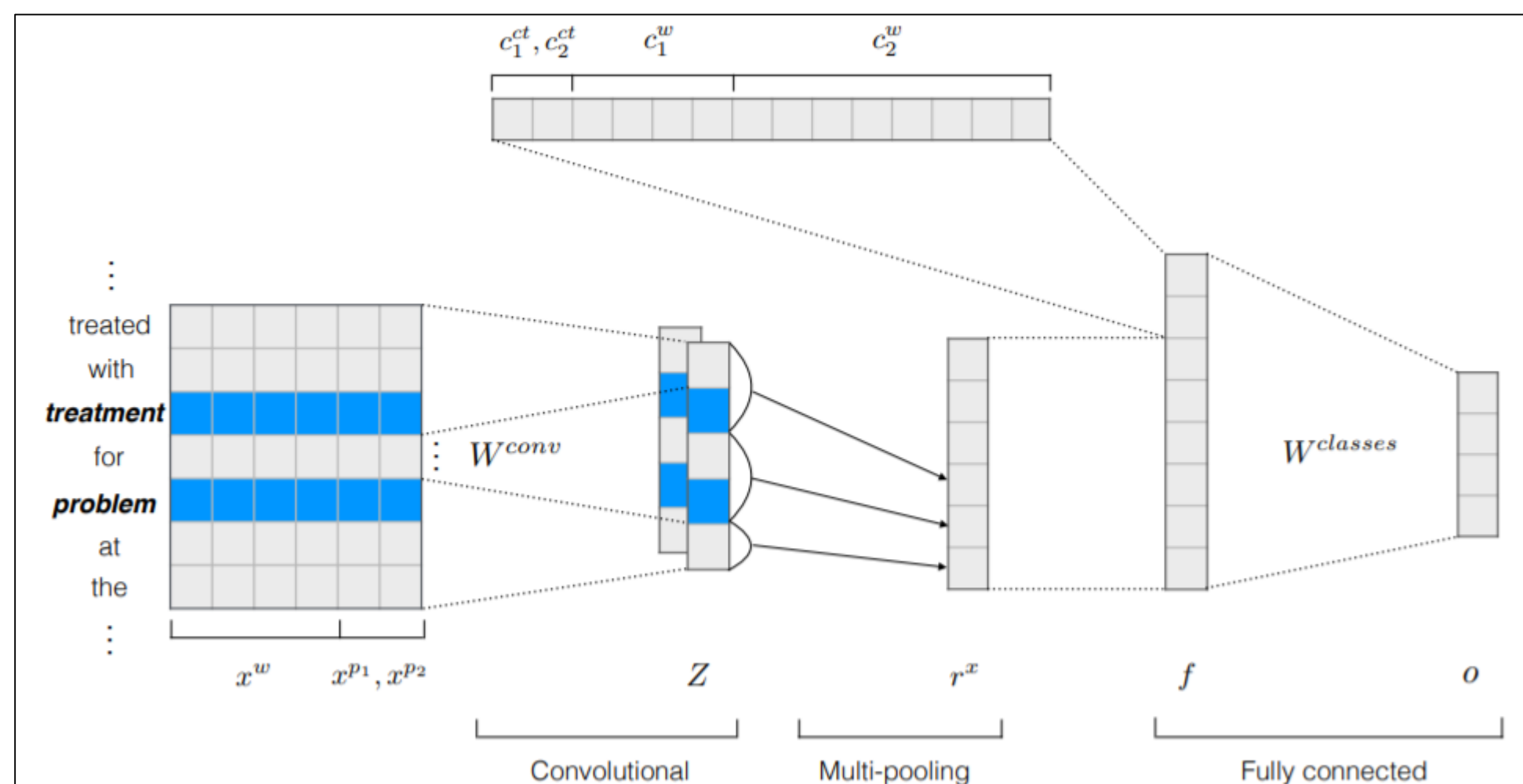


Su, Jia, et al. "Extraction of risk factors for cardiovascular diseases from Chinese electronic medical records." *Computer methods and programs in biomedicine* 172 (2019): 1-10.

Sun, Cong, et al. "Biomedical named entity recognition using BERT in the machine reading comprehension framework." *Journal of Biomedical Informatics* 118 (2021): 103799.

医疗知识图谱

关系抽取



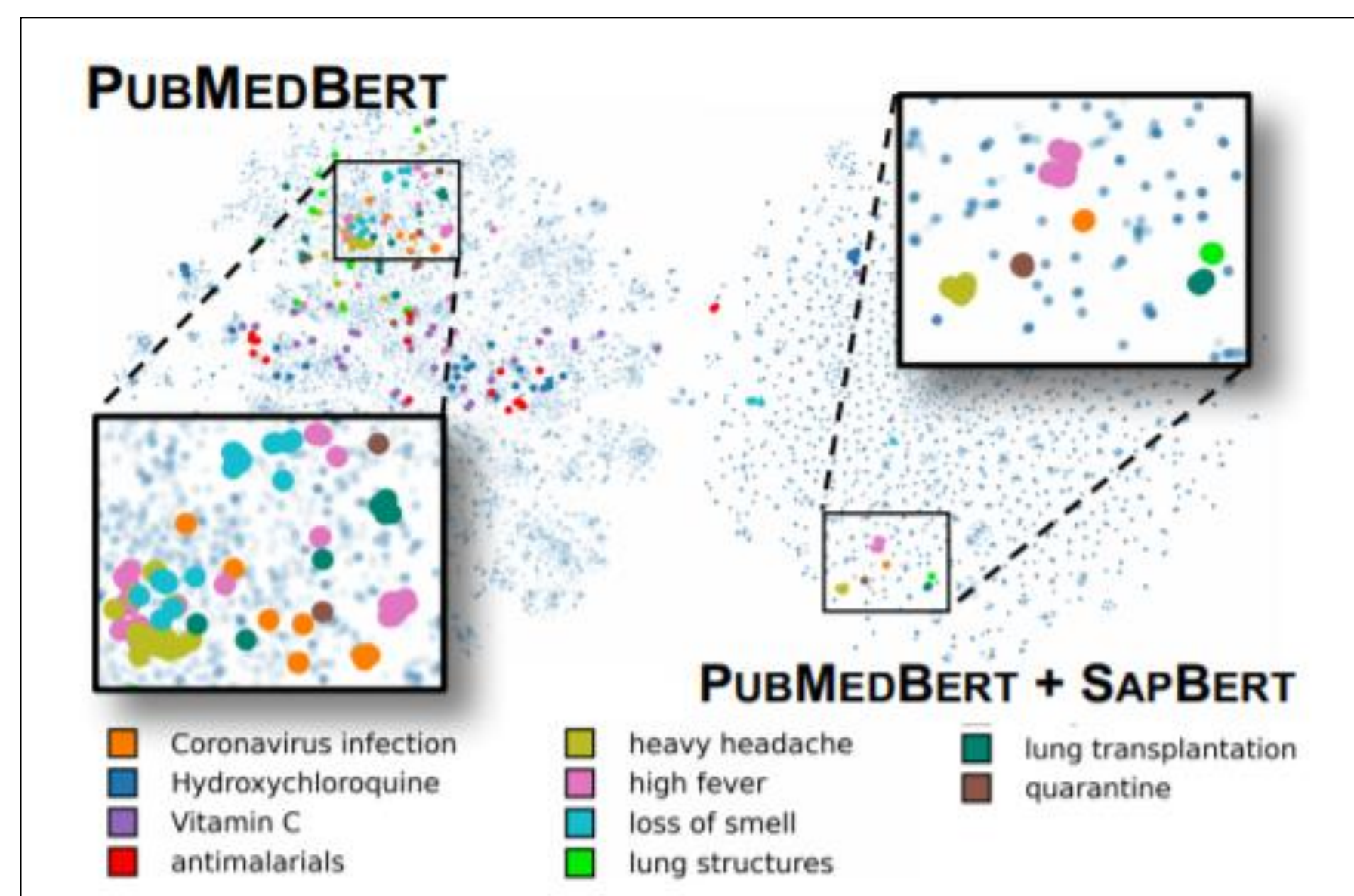
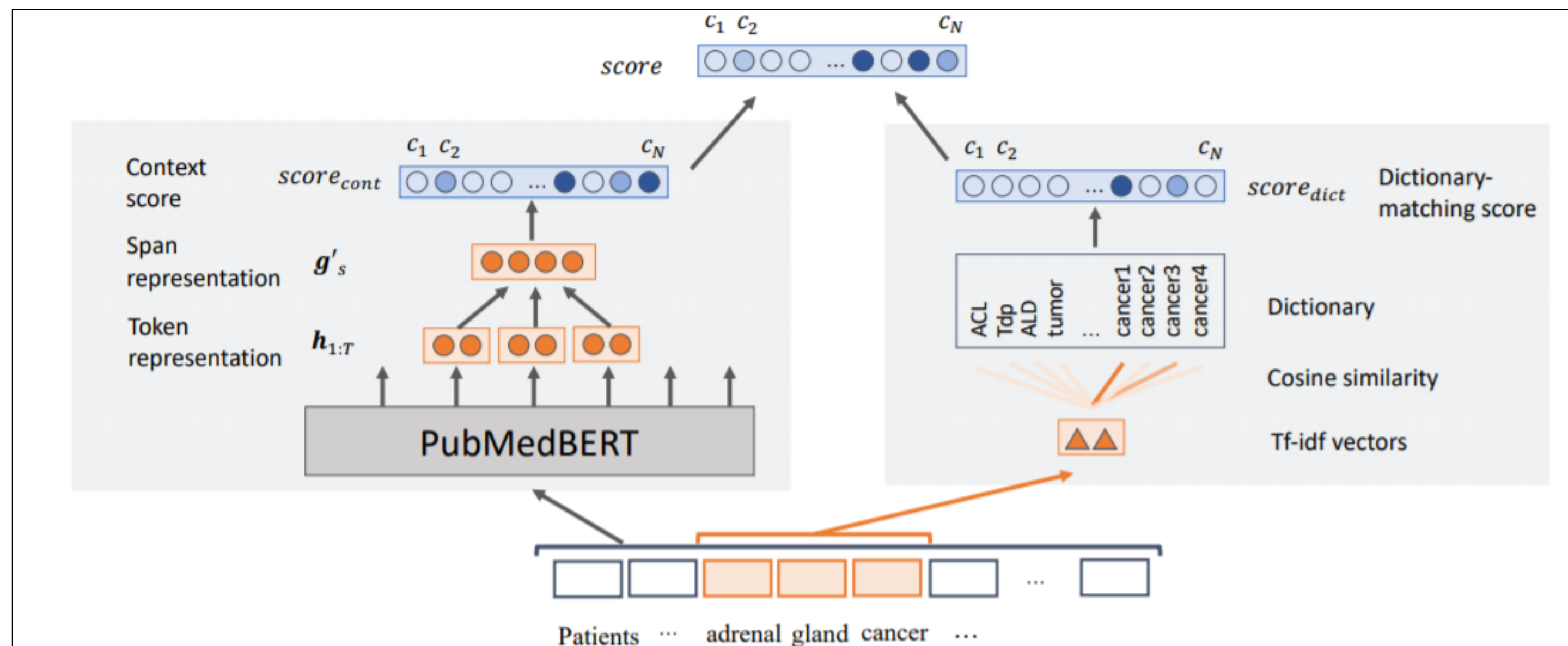
He, Bin, Yi Guan, and Rui Dai. "Classifying medical relations in clinical text via convolutional neural networks." *Artificial intelligence in medicine* 93 (2019): 43-49.

Li, Fei, et al. "A neural joint model for entity and relation extraction from biomedical text." *BMC bioinformatics* 18.1 (2017): 1-11.

Luo, Ling, et al. "A neural network-based joint learning approach for biomedical entity and relation extraction from biomedical literature." *Journal of biomedical informatics* 103 (2020): 103384.

医疗知识图谱

知识对齐

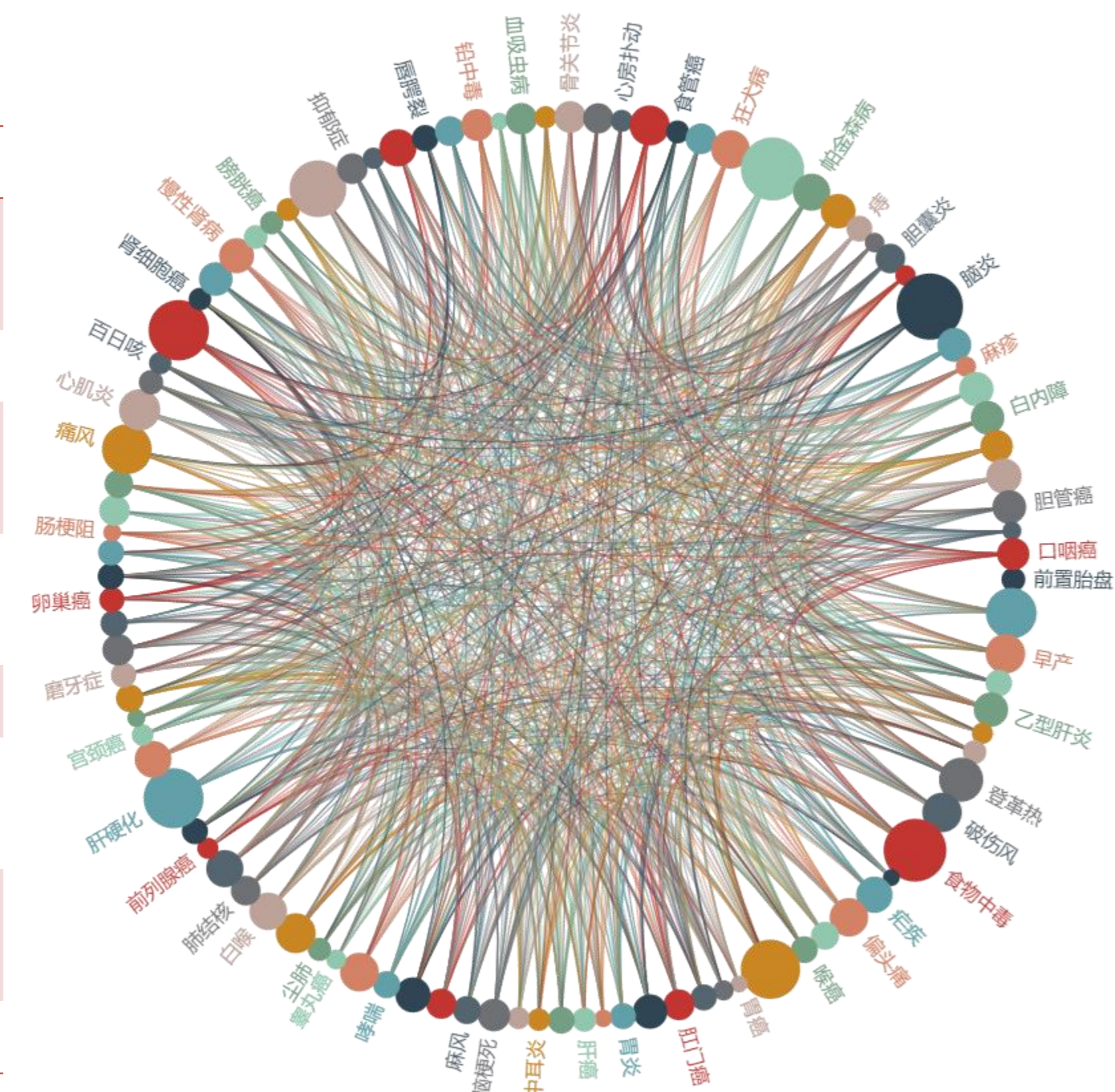


Ujiie, Shogo, et al. "End-to-end Biomedical Entity Linking with Span-based Dictionary Matching." *arXiv preprint arXiv:2104.10493* (2021).

Liu, Fangyu, et al. "Self-alignment pretraining for biomedical entity representations." *arXiv preprint arXiv:2010.11784* (2021).

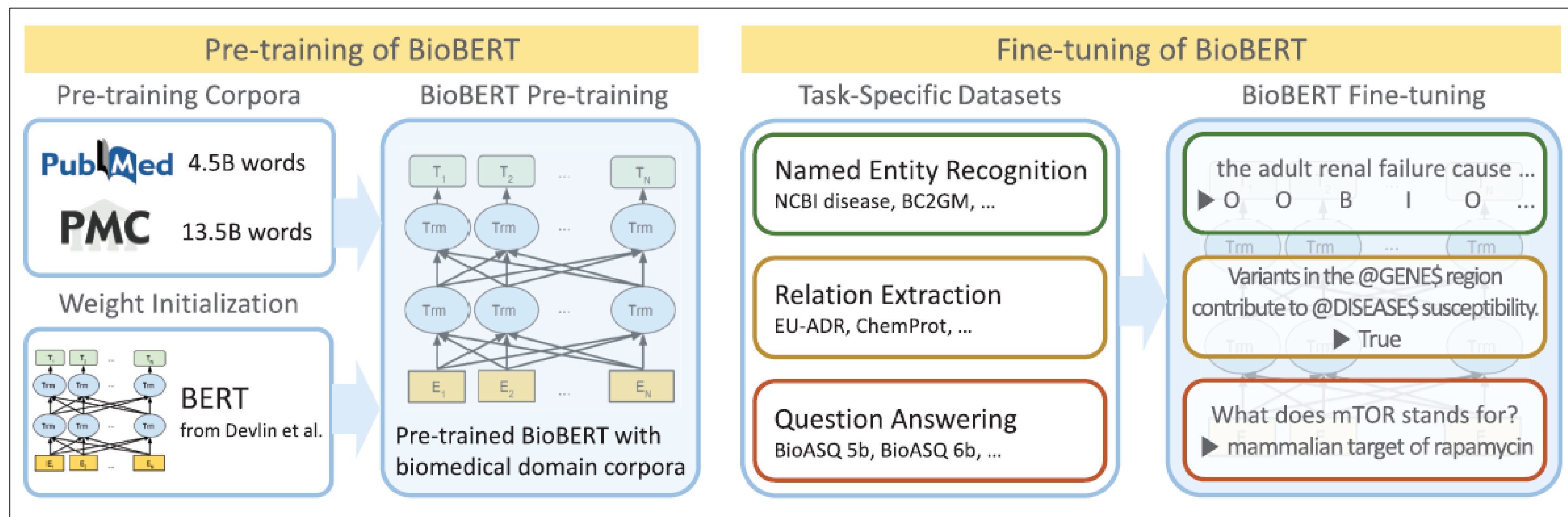
医疗知识图谱

语言	标题及作者	规模	本体
中文	CMeKG 北大计算语言所/郑州大学/鹏城实验室	100W三元组	医疗通用实体及关系
英文	OMAHA医疗知识图谱 OMAHA联盟	40W三元组	医疗通用实体及关系
	新冠基本信息图谱 (华为云/浙大陈华钧团队2020)	190W三元组	基因与蛋白关系
	QMKG高质量医疗图谱 (Goodwin and Harabagiu 2013)	-	通用医疗实体及关系
	I2b2系列评测构建语料	-	通用、药物、时间等实体及关系
	DRKG新冠药物图谱(loannidis et al 2020)	580W三元组	基因、药物、副作用等实体及关系
	BioKG生物医疗图谱 (Walsh et al 2020)	200W三元组	蛋白质、药物、疾病等实体及关系
	CKG临床组学知识图谱 (Mann Labs)	217W三元组	通用，基因组学等实体及关系



医疗预训练模型

BioBERT



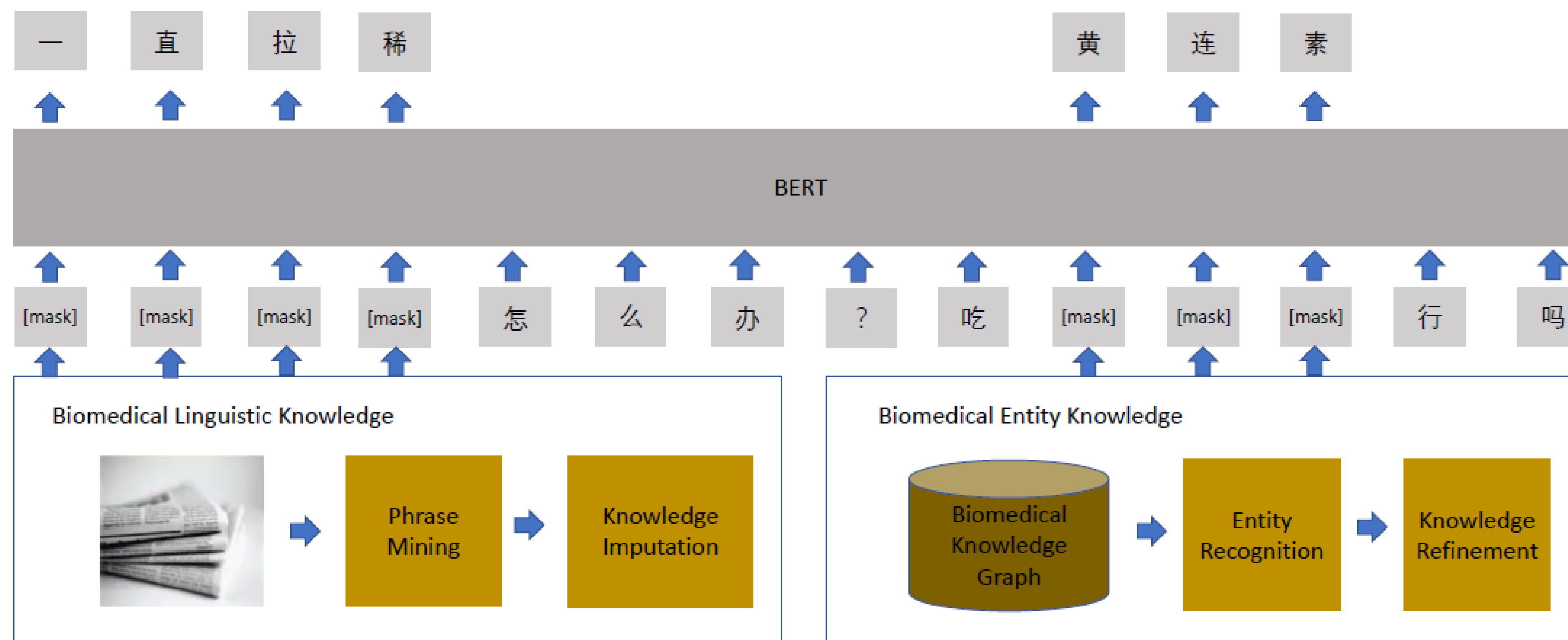
医疗预训练模型相较于通用模型在生物医学文本挖掘任务如实体识别、关系抽取、问答等有着明显的优势

在公开数据集上，NER在F1值上提高2个点，关系抽取F1值提高2.8个点，QA任务上MRR提高7.6个点

Lee, Jinhyuk, et al. "BioBERT: a pre-trained biomedical language representation model for biomedical text mining." Bioinformatics 36.4 (2020): 1234-1240.

医疗预训练模型

MC-BERT



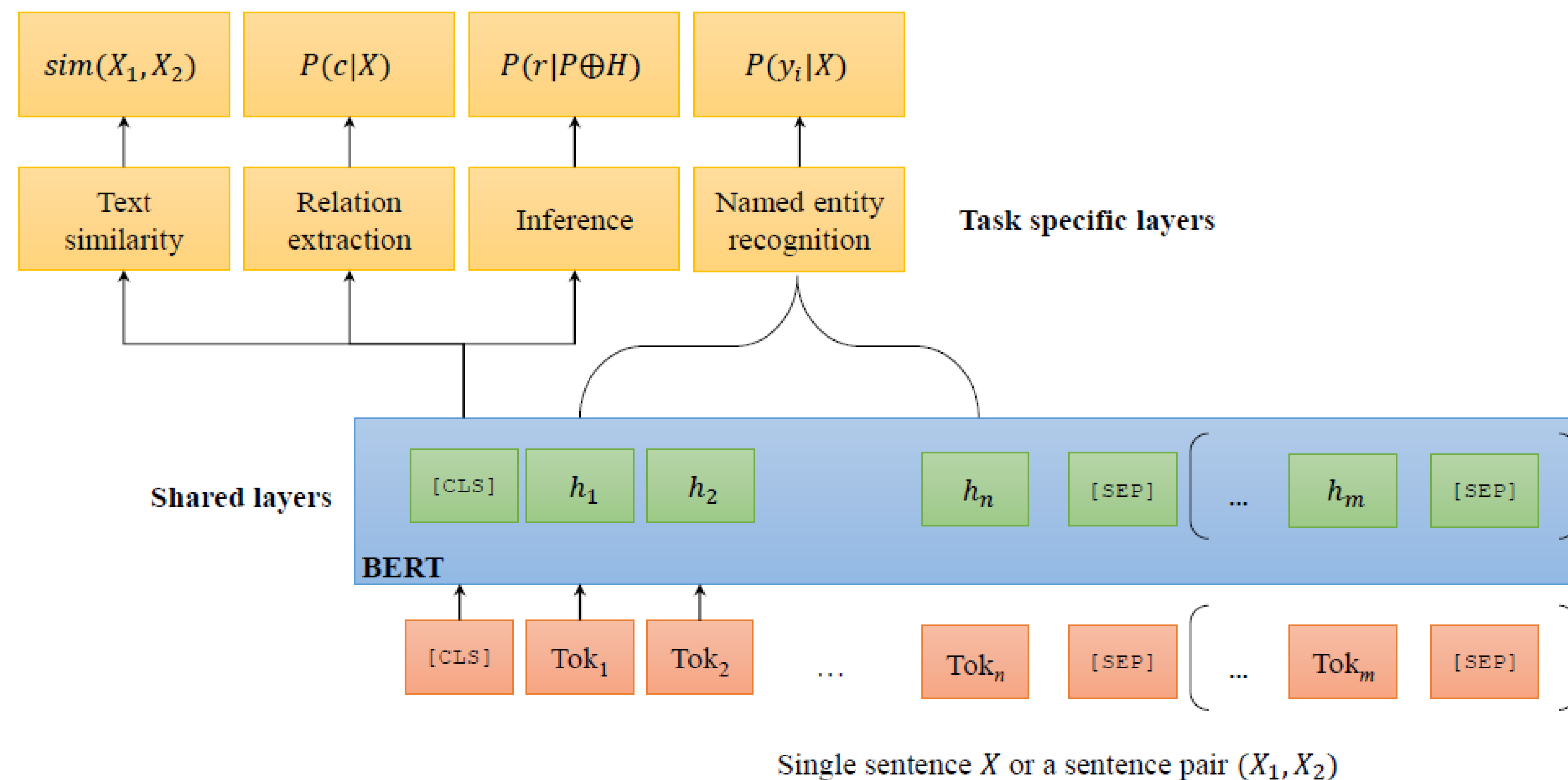
医学实体全遮盖

医学短语全遮盖

Ningyu Zhang, Qianghui Jia, Kangping Yin, Liang Dong, Feng Gao, and Nengwei Hua. 2020. Conceptualized Representation Learning for Chinese Biomedical Text Mining. In WSDM ' 20: , February 3–7, 2020, Houston. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

医疗预训练模型

MT-BERT

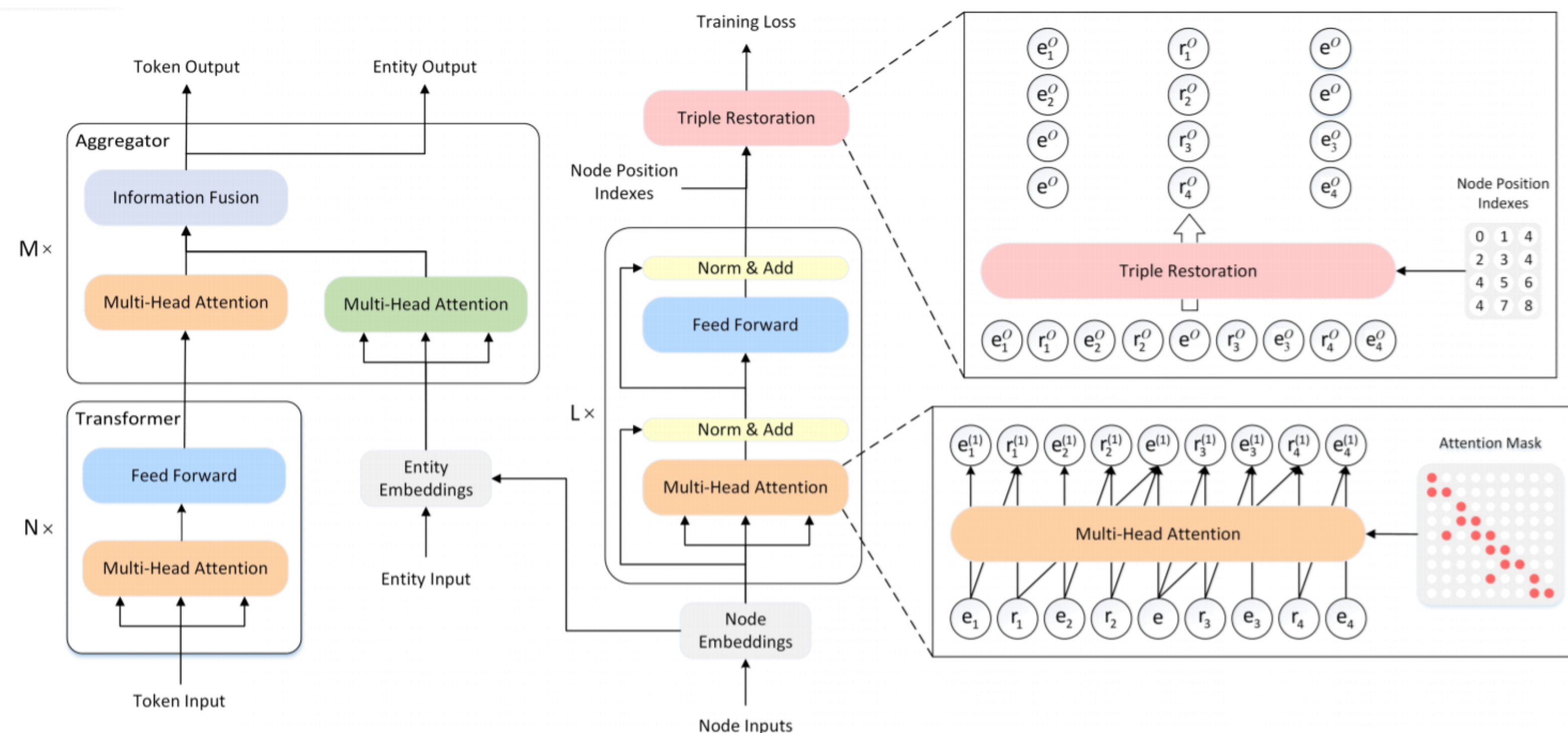


结合下游任务文本相似、关系抽取、推理、NER的多任务学习
共享BERT参数

Peng, Yifan, Qingyu Chen, and Zhiyong Lu. "An empirical study of multi-task learning on BERT for biomedical text mining." arXiv preprint arXiv:2005.02799 (2020).

医疗预训练模型

BERT-MK



通过知识图谱三元组学习实体表示

整合实体表示到PTM训练，将知识图谱内的信息融合到预训练模型中，指导模型参数学习

He, Bin, et al. "Integrating graph contextualized knowledge into pre-trained language models." *Findings of the EMNLP 2020*, p2281-2290.

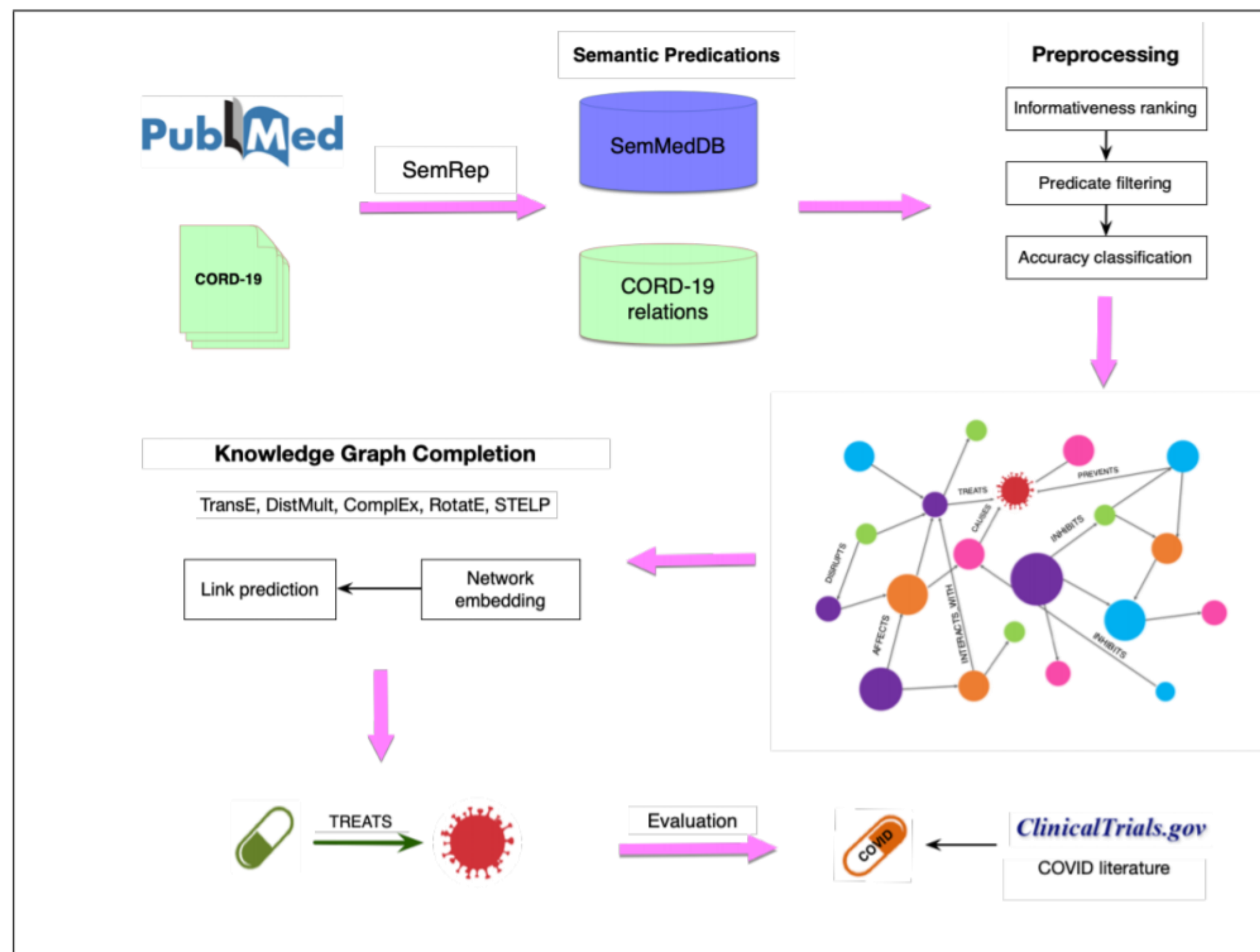
04 融合领域知识推 理技术

医疗知识计算

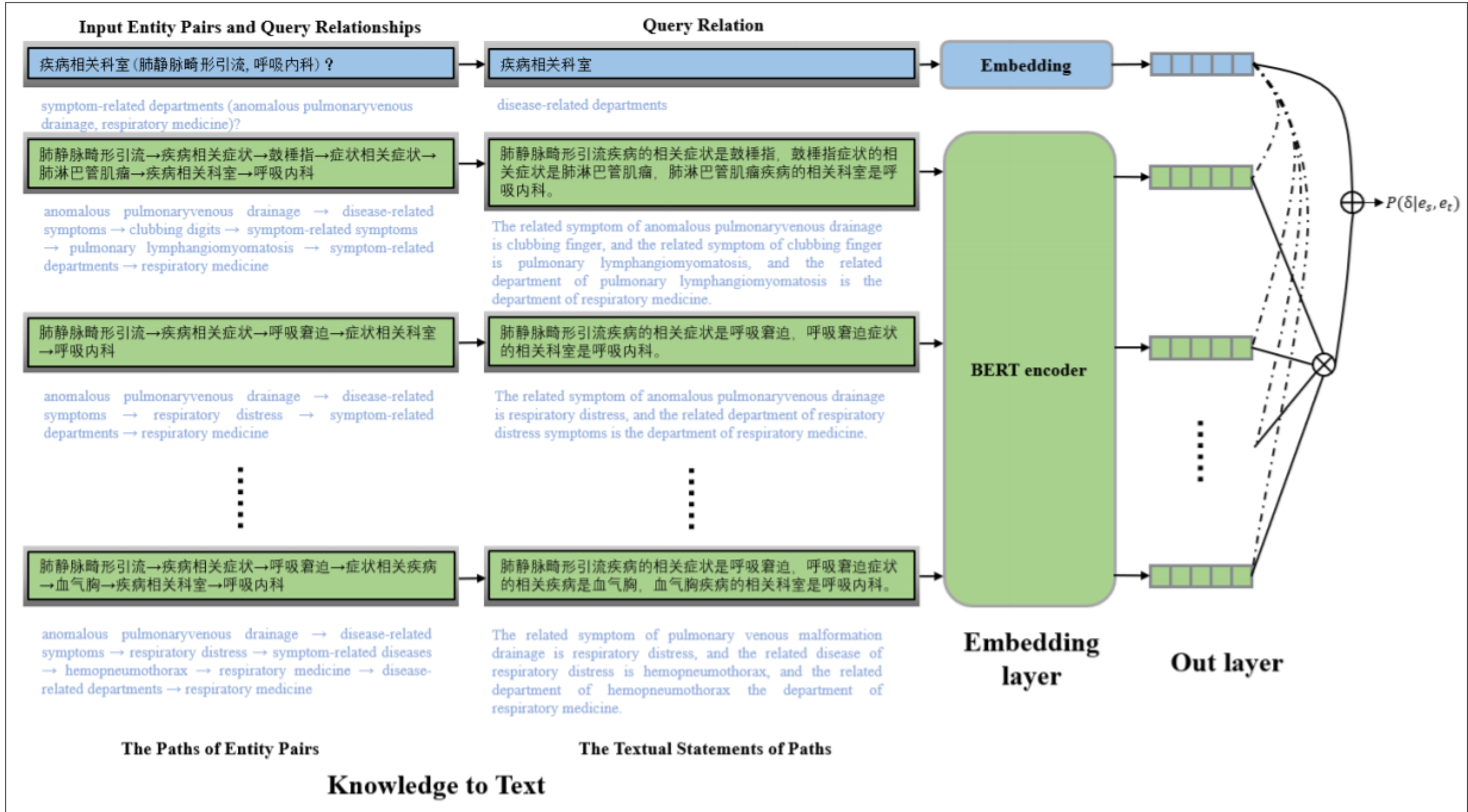
医疗知识计算

在AI中“推理”通常包括两方面：

✓ Reasoning：生成新知识，要求有完整证据链，具有可解释性



医疗知识计算



Lan, Yinyu, et al. "Path-based knowledge reasoning with textual semantic information for medical knowledge graph completion." *arXiv preprint arXiv:2105.13074* (2021).

医疗知识计算

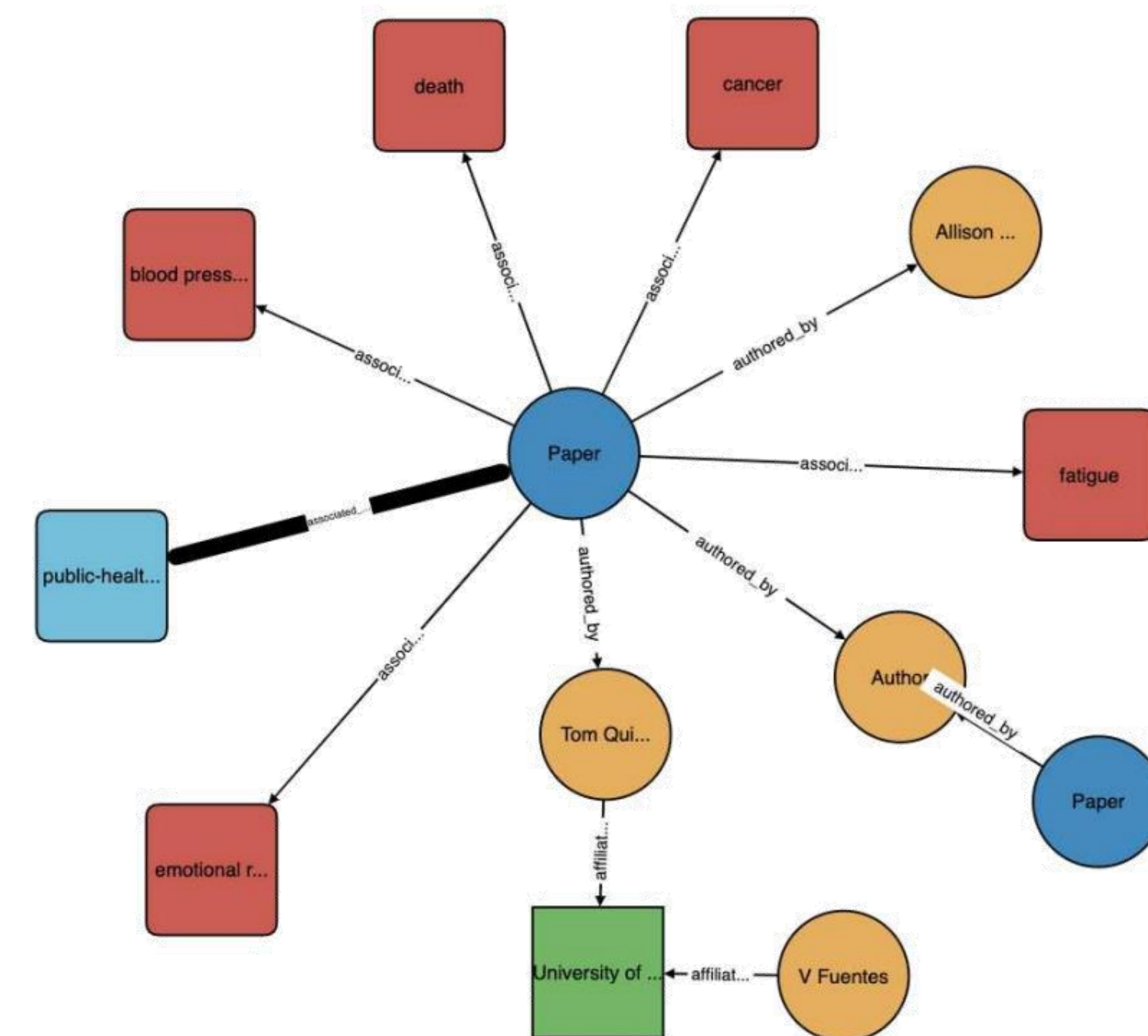
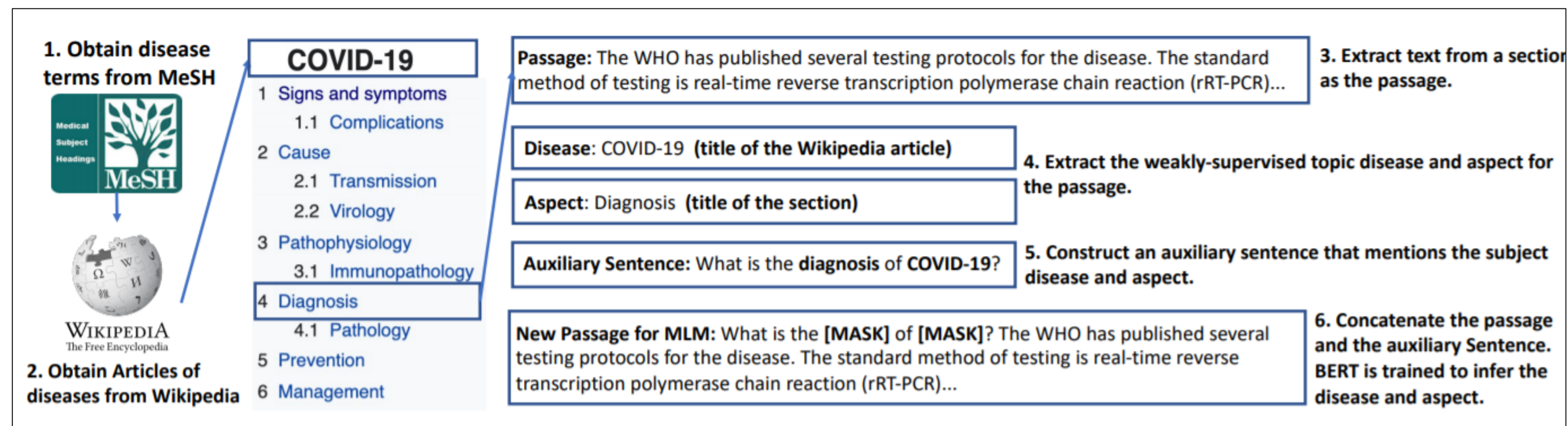
✓ Inference: 一般指训练模型结束后的预测

排序类问题

问答

医疗文献检索

推荐

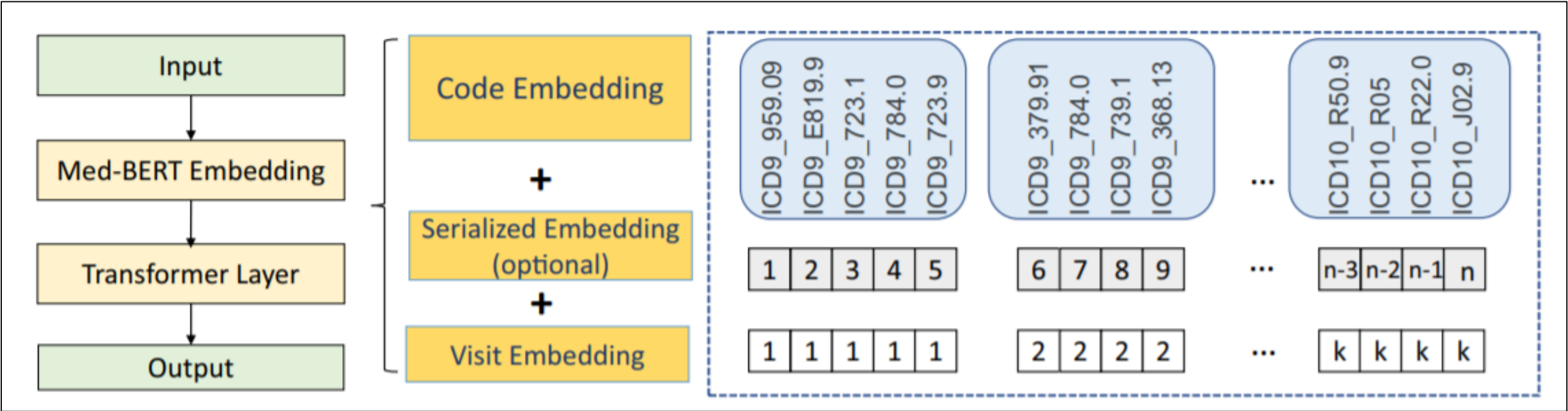
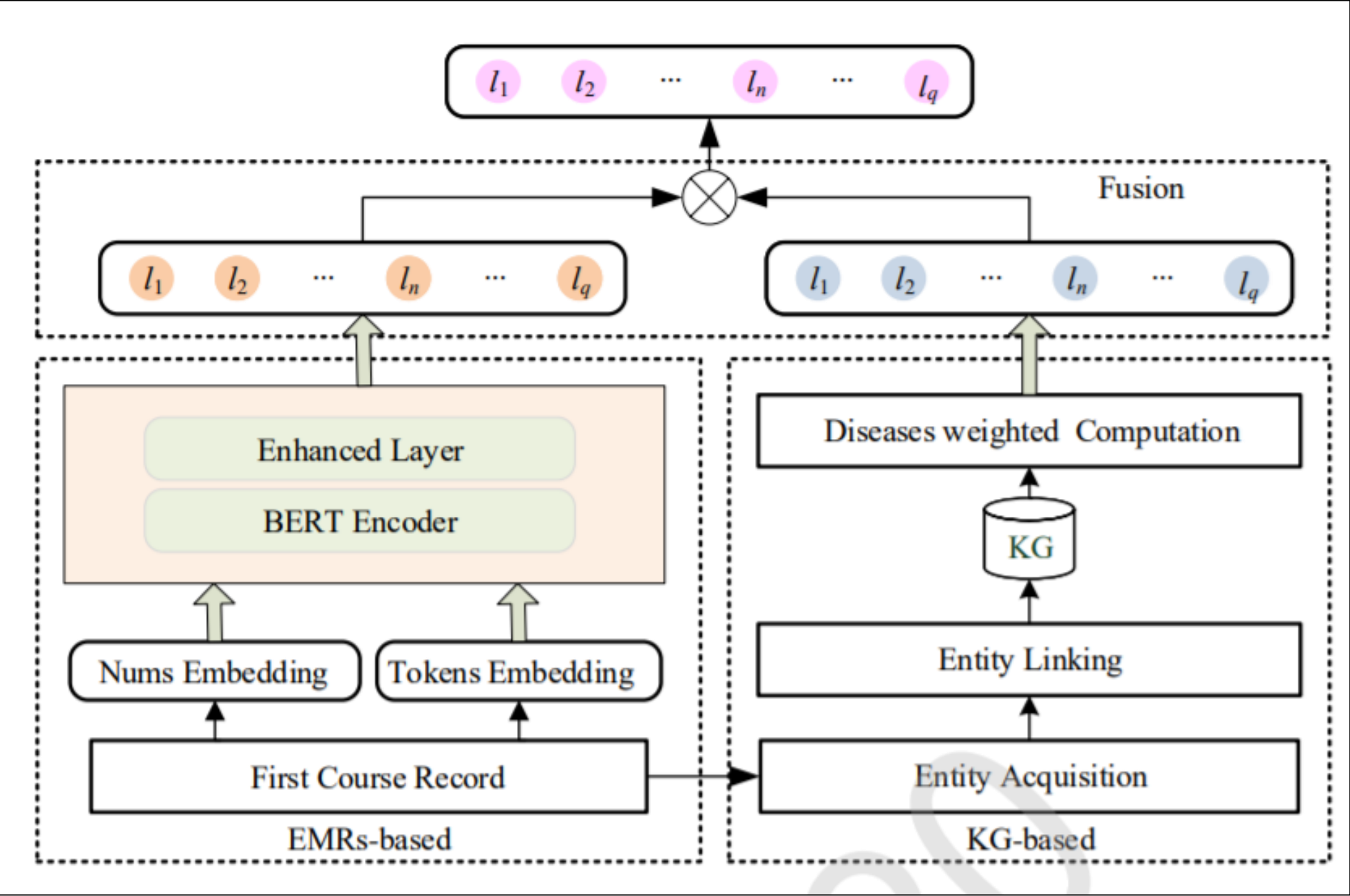


He, Yun, et al. "Infusing disease knowledge into BERT for health question answering, medical inference and disease name recognition." arXiv preprint arXiv:2010.03746 (2020).

Wise, Colby, et al. "COVID-19 knowledge graph: accelerating information retrieval and discovery for scientific literature." arXiv preprint arXiv:2007.12731 (2020).

医疗知识计算

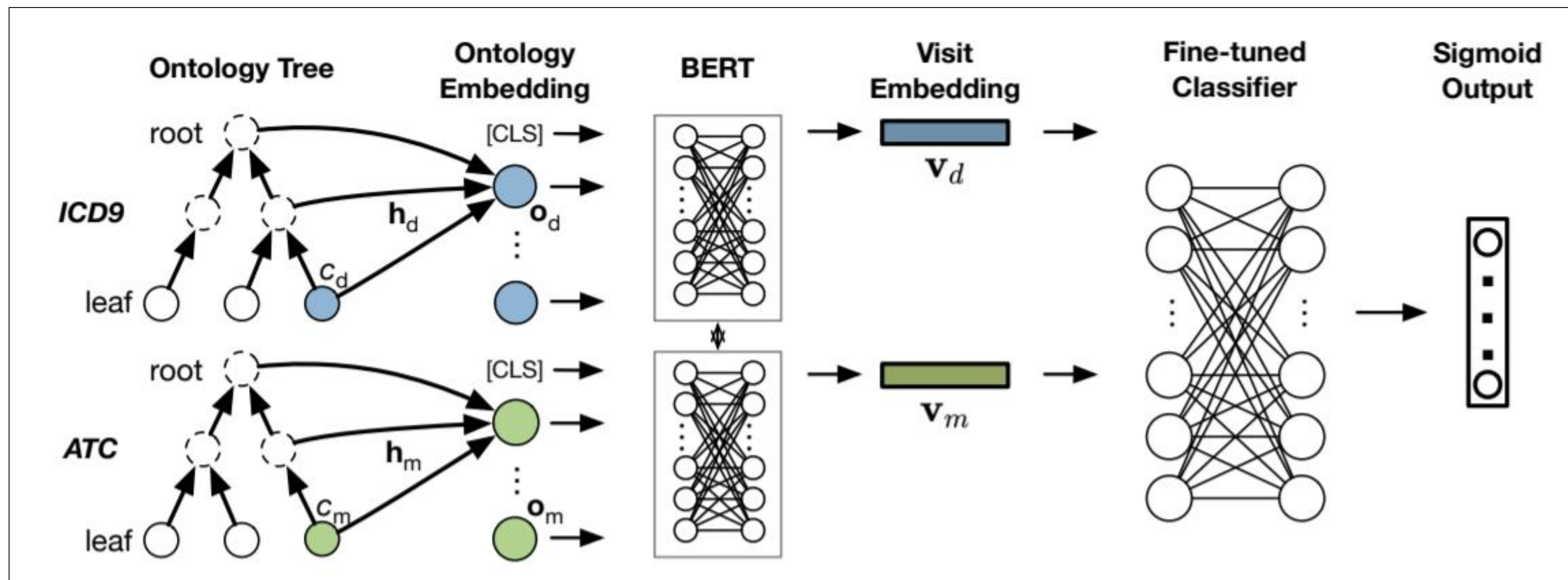
分类问题
疾病诊断
疾病预测



Zhang, Kunli, et al. "Knowledge-Enabled Diagnosis Assistant Based on Obstetric EMRs and Knowledge Graph." *China National Conference on Chinese Computational Linguistics*. Springer, Cham, 2020.
Rasmy, Laila, et al. "Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction." *NPJ digital medicine* 4.1 (2021): 1-13.

医疗知识计算

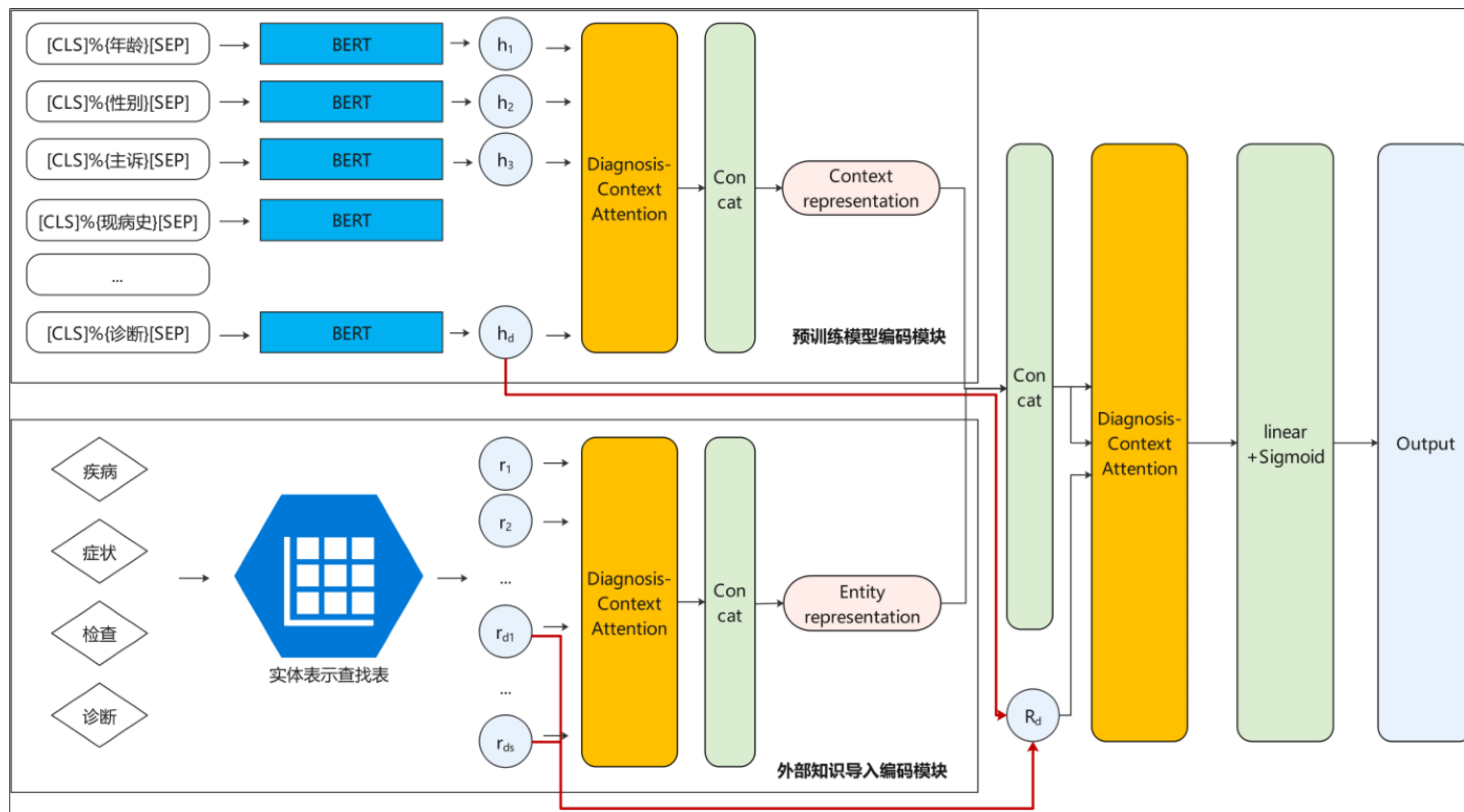
分类问题
治疗建议



Shang, Junyuan, et al. "Pre-training of graph augmented transformers for medication recommendation." *arXiv preprint arXiv:1906.00346* (2019).

医疗知识计算

临床决策任务通用框架



一种整合预训练模型和知识图谱做推理任务的框架

预训练模型对EHR等患者个人信息进行语义编码

知识图谱对患者的关键实体如症状、疾病、检查、诊断等进行整合，同和融入外部知识

医疗知识计算

临床决策支持任务实验

治疗推荐	P	R	F1
Bert(100M)	0.771	0.793	0.782
Domain-Bert(100M)	0.821	0.816	0.819
疾病诊断			
Bert(100M)	0.856	0.775	0.813
Domain-Bert(100M)	0.863	0.795	0.828



THANKS!

今天的分享就到这里...

欢迎大家微信搜索关注“华为云”

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

