

# 深度生成式模型在临床术语标准化中的应用

北京中科凡语科技有限公司

中国科学院自动化研究所 | 模式识别国家重点实验室 | 自然语言处理团队

汇报人：薛晨

# 目录

CONTENTS

01

企业简介

02

任务分析

03

解决方案

04

实验结果

05

总结与展望

# 企业简介

---

01

## 北京中科凡语科技有限公司

中科院自动化所模式识别国家重点实验室自然语言处理研究团队创立  
专注于人工智能、自然语言处理，破冰于医疗航空教育领域。

- 某大型三甲医院：电子病历结构化、医疗知识图谱、辅助决策支持系统
- 知名互联网企业：基于医疗健康数据与主诉的分诊和问诊的人工智能技术开发
- 知名药企：对话机器人、医疗翻译引擎

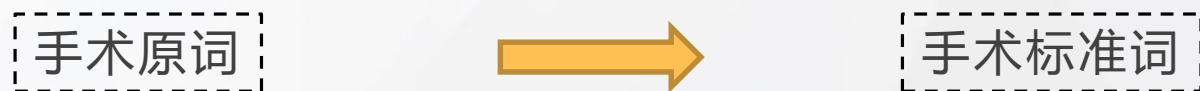
# 任务分析

---

02

## 任务说明

中文电子病历中挖掘出的真实手术实体进行语义标准化。



手术原词	手术标准词	特点
微创经椎间孔下腰4/5椎间盘摘除术	腰椎间盘突出术	原词书写自由，信息冗余
腺样体消融术	腺样增殖体切除术不伴扁桃体切除术	原词书写简略，信息隐藏
亲体肾移植术	肾异体移植术##与供者有血缘关系的活体移植	一个原词可能对应多个标准词
(输尿管（或尿道）镜右侧)输尿管镜下取石术+输尿管镜下钬激光碎石术	经尿道输尿管/肾盂激光碎石取石术	包含多种术式信息的原词不一定对应相同数量的原词

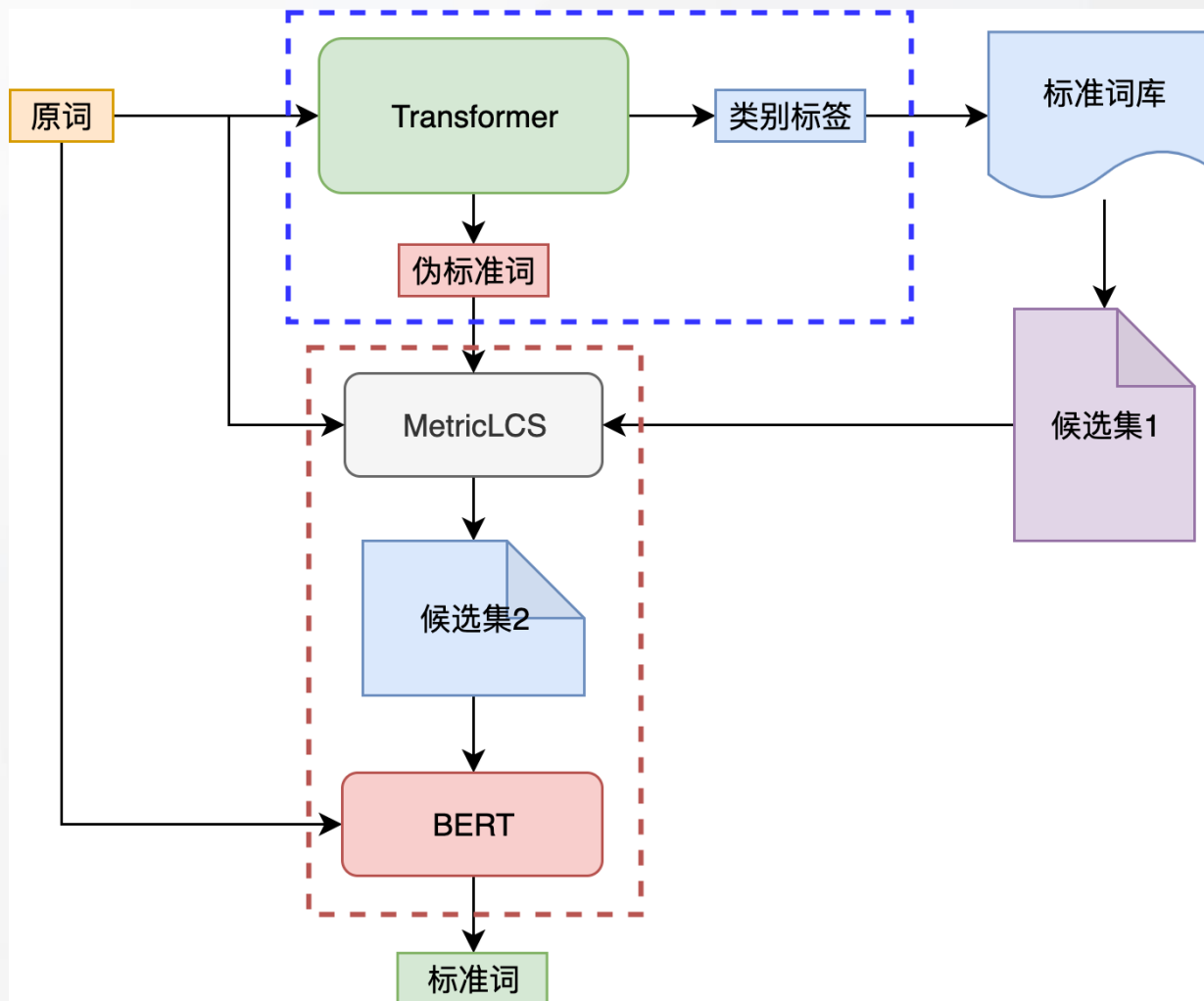
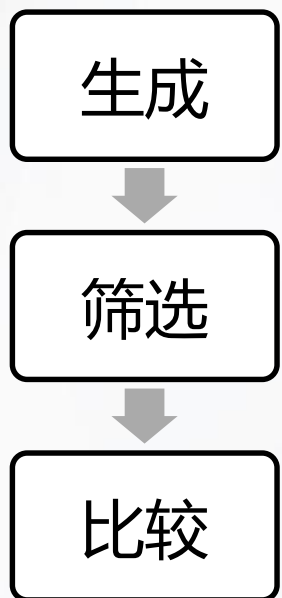
类别	数量	细分类	细分类数量
原词切分数 = 标准词切分数	115	切分数为2	92
		切分数为3	16
		切分数为4	7
原词切分数 < 标准词切分数	80	比例为1: 2	55
		比例为1: 3	9
		比例为1: 4	4
		比例为2: 3	7
		比例为3: 4	4
		比例为2: 4	1
原词切分数 > 标准词切分数	134	比例为2: 1	125
		比例为3: 1	5
		比例为3: 2	1
		比例为4: 3	2
		比例为9: 7	1

# 解决方案

---

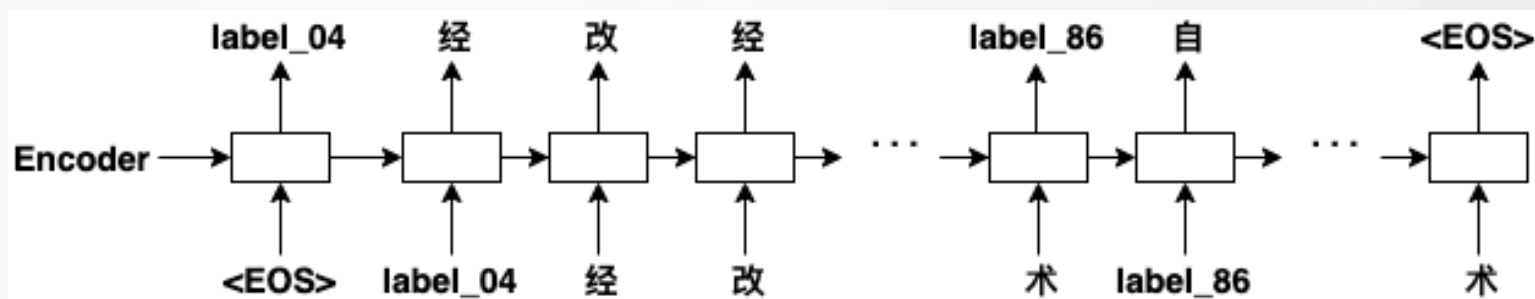
03





source:

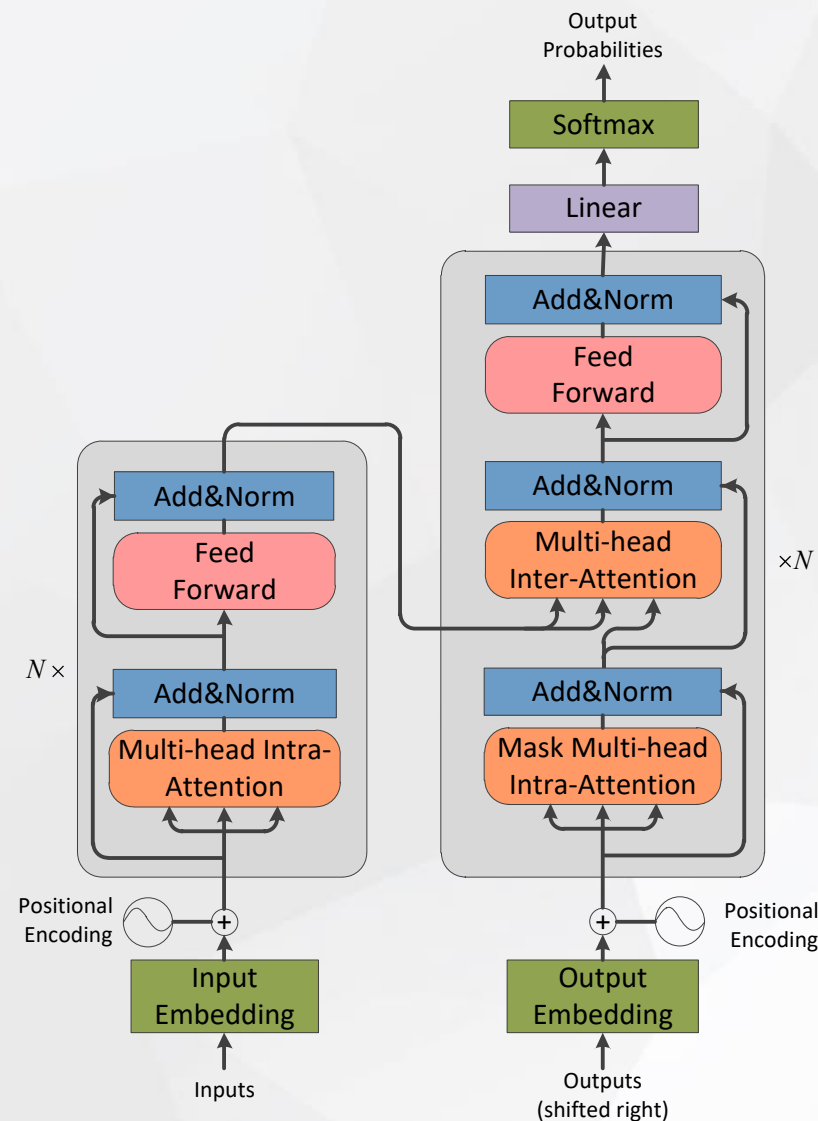
经改良扩大迷路径路左侧听神经瘤摘除+腹部取脂术

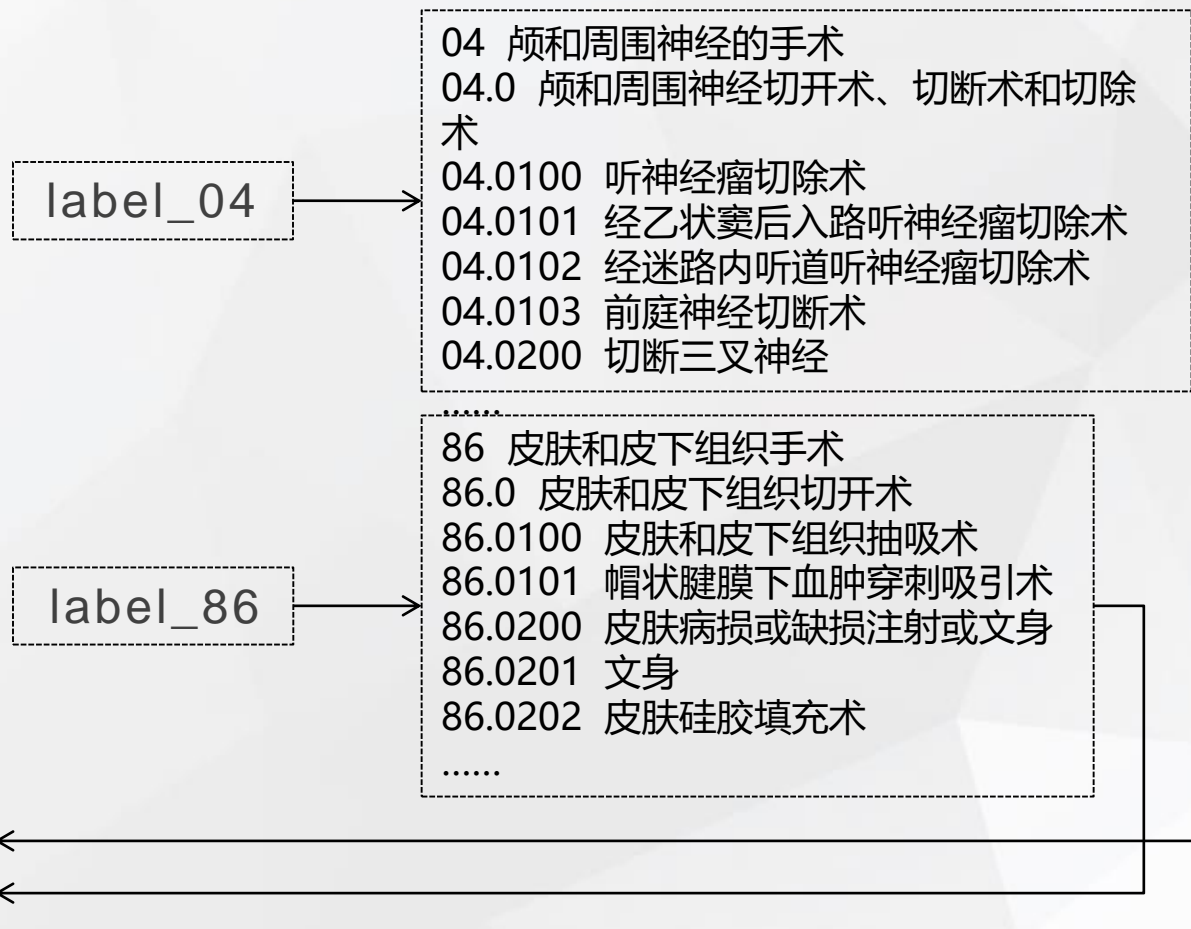
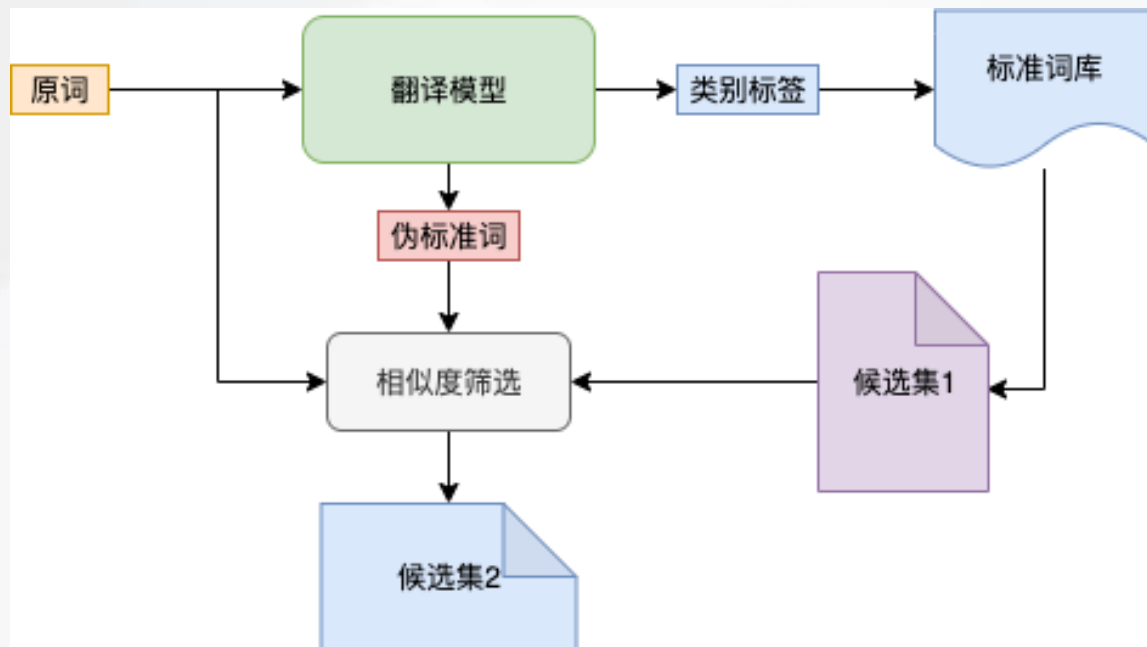


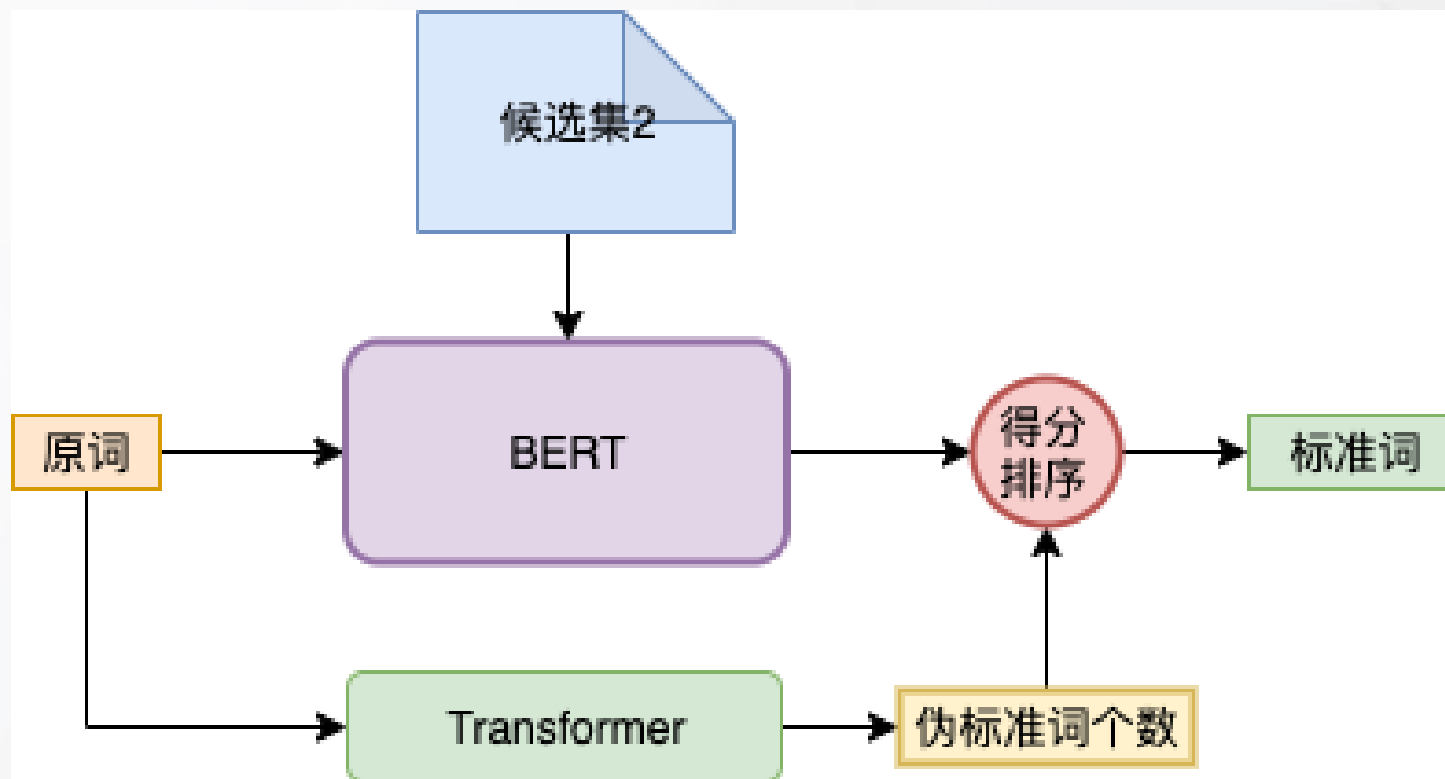
target:

label\_04 经迷路内听道听神经瘤切除

label\_86 自体脂肪移植术







# 实验结果

---

04

字到字模型:

source: 右 腮 腺 全 切 除 术

target: label\_26 腮 腺 切 除 术

词到词模型:

source: 右 腮腺 全 切除 术

target: label\_26 腮腺 切除 术

字到词模型:

source: 右 腮 腺 全 切 除 术

target: label\_26 腮腺 切除 术

	伪标准词	类别标签	伪标准词数目
字到字	<b>84.7%</b>	94.9%	96.7%
字到词	82.9%	<b>95.1%</b>	<b>97.3%</b>
词到词	78.9%	93.3%	97.1%

dev	test
89.4%	91.3%

# 总结与展望

---

05



## 模型优势

- ✓ 本模型没有使用额外标注数据
- ✓ 本模型没有进行任何人工后处理规则添加

## 未来优化

- 引入分词词典（部位、术式、入路.....）
- 引入专家知识





# THANK YOU

北京中科凡语科技有限公司  
中国科学院自动化研究所 | 模式识别国家重点实验室 | 自然语言处理团队