

# 基于检索和重排序的手术 名称标准化研究

陈漠沙，谭传奇，仇伟

阿里巴巴

2019/11/24



Define Smarter  
Tomorrow.

# 大纲

- 任务介绍
- 算法
  - 检索算法
  - 重排序算法
  - 集成结果和后处理

# 临床术语标准化

- 临床术语标准化任务是医学统计中不可或缺的一项任务
  - 诊断、手术、药品、检查、化验、症状
- 可作为语义相似度匹配任务
- 术语标准化后，才可对电子病历进行后续的统计分析

# 手术名称标准化

- 针对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
  - 手术标准词基于《ICD9-2017协和临床版》

# 数据分析

- 数据集中“手术原词”和“归一化标准词”之间存在4种匹配对应关系：
  - “一对一”关系：一个手术原词对应一个归一化标准词
  - “一对多”关系：一个手术原词对应一个多个归一化标准词
  - “多对一”关系：多个手术原词对应一个归一化标准词
  - “多对多”关系：多个手术原词对应多个归一化标准词

手术原词	归一化标准词
右肾上腺巨大肿瘤切除术	肾上腺病损切除术
右眼硅油取出联合人工晶体II期植入术	玻璃体硅油取出术 + 眼内人工晶状体二期置入
右叶甲状腺切除 + 左叶甲状腺部分切除术	单侧甲状腺切除伴他叶部分切除术
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



# 数据分析

- 1个手术原词最多可以对应到7个标准词
- 约10%的数据非“一一对应”

	训练集	验证集	测试集
数据条数	4,000	1,000	2,000
最大长度	144	120	93
平均长度	12.33	12.31	12.35
标准词平均个数	1.07	1.06	1.06
标准词最大个数	7	4	5
“一一对应”比例	0.918	0.913	0.927

# 评测指标

- 对于一条手术原词，其对应N个归一化标准词。模型预测输出了M个归一化标准词，则该条数据的得分为：

$$• S_i = \frac{Count(N \cap M)}{Max(Count(N), Count(M))}$$

- 总得分为：

$$• S = \frac{1}{k} \sum_1^k S_i$$

# 算法

- 候选答案生成
  - 依据手术原词，召回候选答案，尽可能保证召回率
- 候选答案打分
  - 对于召回的候选答案，进行打分排序
- 集成结果和后处理





编码文件	
ID	标准词
00	操作和介入NEC
00.0	治疗性超声
00.0100	头和颈部血管治疗性超声

标注文件	
手术原词	标准词
横结肠造口还纳术	横结肠造口闭合术
右肾上腺巨大肿瘤切除术	肾上腺病损切除术
左侧单侧乳房根治性切除术	头单侧根治性乳房切除术

建立索引



椎体成形术 查询索引

候选答案

检索编码文件

ID	候选标准词
78.4904	椎骨成形术
81.6601	经皮椎体球囊扩张成形术
81.6500	经皮椎骨成形术
03.0913	椎间盘镜下椎管成形术
.....	.....

检索标注文件

手术原词	候选标准词
经皮椎体成型术	经皮椎骨成形术
椎体后凸成型术	经皮椎体增强
经皮球囊扩张椎体后突成形术	经皮椎体球囊扩张成形术
.....	.....

BERT打分

BERT输入

[CLS]椎体成形术[SEP]椎骨成形术[SEP]  
[CLS]椎体成形术[SEP]经皮椎骨成形术[SEP]  
[CLS]椎体成形术[SEP]经皮椎体增强[SEP]

输出标准词

经皮椎骨成形术

# 候选答案生成

- 使用Lucene工具生成候选答案
  - 其默认排序基于TF-IDF和向量空间模型
- 两种索引
  - “编码—标准词”：查找与待归一化的“手术原词”最相近的“标准词”
  - “标注历史”：在标注数据上查找与待归一化的“手术原词”最相近的数据，并取该条数据的归一化“标准词”作为候选答案

Top	1	2	5	10	20	50
编码检索	49.5	57.4	68.8	77.4	86.0	91.7
历史检索	71.1	78.3	83.1	85.9	88.1	88.7
编码检索 + 历史检索	86.8	91.5	95.4	97.8	99.2	99.6

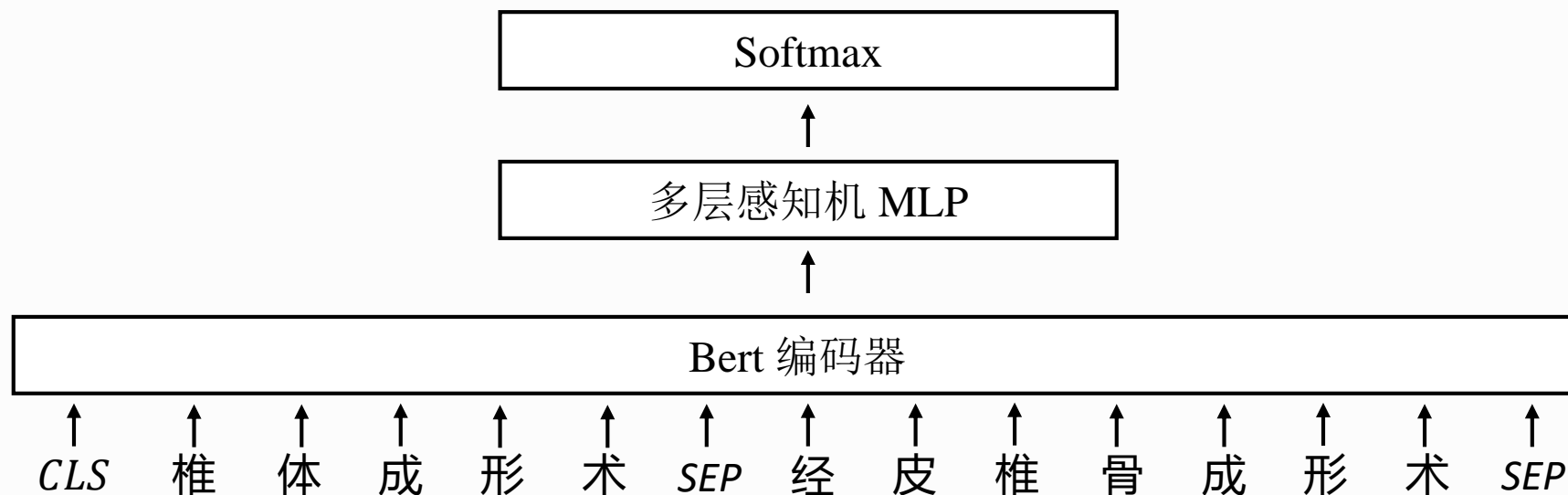
# 候选答案生成

- 未召回例子分析

手术原词	归一化标准词
左侧化疗泵置入+右侧脑室腹腔分流术	颅腔或组织的导管置入术
LC	腹腔镜下胆囊切除术
原发性肝癌HIFU术	肝病损超声刀治疗
化疗泵置入+脑室腹腔分流术	颅腔或组织的导管置入术

# BERT重排序

- 任务定义：某一个归一化标准词是否被包含在手术原词之中
- 即将整个手术原词和一个归一化标准词作为一条输入，判别该归一化标准词是否应该出现在最终输出之中



# BERT重排序

- 选取“BERT - 取Top20候选 + 10倍正例”的设定

模型	准确率
BERT - 取Top10候选	85.6
BERT - 取Top20候选	87.5
<b>BERT - 取Top20候选 + 10倍正例</b>	<b>89.1</b>
BERT - 取Top20候选 + 正确答案同子类的其它条目作为负例	88.6
RoBERTa - 取Top20候选 + 10倍正例	87.4

# 后处理

- 采用了两种后处理方式
  - 若原词中不包含“+”号，则只输出得分最高的一个手术标准词 (+1%)
  - 若所有候选得分均未超过选定的阈值，则从候选标准词中输出得分最高的一个手术标准词 (+2.5%)

# 集成模型

- 将验证集均分3份，每次取其中2份加入训练集，剩余1份做验证

序号	训练数据	验证数据	BERT
1	训练集	验证集（全）	89.1
2	训练集 + 验证集（1， 2）	验证集（3）	89.0
3	训练集 + 验证集（1， 2）	验证集（3）	88.6
4	训练集 + 验证集（1， 2）	验证集（3）	89.4
5	训练集 + 验证集（1， 3）	验证集（2）	88.6
6	训练集 + 验证集（1， 3）	验证集（2）	88.7
7	训练集 + 验证集（1， 3）	验证集（2）	89.4
8	训练集 + 验证集（2， 3）	验证集（1）	88.7
9	训练集 + 验证集（2， 3）	验证集（1）	88.9

# 集成方式

- 投票
  - 取10个模型中票数多的作为输出
  - 票数并列最高则都输出
- 最终结果
  - 测试集： 92.8



# 谢谢

基于检索和重排序的手术名称标准化研究

陈漠沙，谭传奇，仇伟

阿里巴巴

2019/11/24



Define Smarter  
Tomorrow.