

学期汇报

高桢 2021/03/17



上学期总结

- 论文
 - 图谱补全
- 项目
 - -专家推荐系统改良&维护(未完待续)
 - 华为知识图谱演化(未完待续)
 - 从文档中抽取关系
 - 张量分解模型做图谱补全
- 比赛
 - SemEval Task9 表格事实验证
 - 模仿SOTA方法(TAPAS by Google)



下学期计划



主要计划

- TAPAS核心思路:
 - 将问题和表格每一行拼接成token序列
 - 每个输入 token 都被表示成其词、绝对位置、 句段、列和行以及数值排序的嵌入之和, 输 入BERT 层

Table					
col1	col2				
0	1				
2	3				

Token Embeddings	[CLS]	query	?	[SEP]	col	##1	col	##2	0	1	2	3
	+	+	+	+	+	+	+	+	+	+	+	+
Position Embeddings	POS	POS ₁	POS ₂	POS ₃	POS ₄	POS ₅	POS ₆	POS ₇	POS ₈	POS ₉	POS ₁₀	POS ₁₁
_	+	+	+	+	+	+	+	+	+	+	+	+
Segment Embeddings	SEG _o	SEG _o	SEG _o	SEG _o	SEG ₁	SEG ₁	SEG ₁	SEG ₁	SEG ₁	SEG ₁	SEG ₁	SEG ₁
_												
	+	+	+	+	+	+	+	+	+	+	+	+
Column Embeddings	COL _o	+ COL₀	+ COL₀	COL _o	COL,	COL ₁	COL ₂	COL ₂	COL ₁	+ COL₂	COL,	COL ₂
Embeddings	COL _o	COL _o	COL _o	COL _o	COL,	COL,	COL ₂	COL ₂	COL,	COL ₂	COL,	COL ₂
	COL _o ROW _o	COL _o + ROW _o	COL _o + ROW _o	COL _o + ROW _o	COL, ROWo	COL, + ROW _o	COL ₂ + ROW ₀	COL ₂ + ROW ₀	COL, + ROW,	COL ₂ + ROW ₁	COL ₁ + ROW ₂	COL ₂ + ROW ₂
Embeddings Row Embeddings	+	+	+	+	COL, ROWo	+	+	+	COL, + ROW, +	COL ₂ + ROW ₁	+	+
Embeddings Row	+	+ ROW _o	+ ROW _o	+	COL, ROW, RANK,	+ ROW _o	+	+ ROW _o	COL, ROW, RANK,		+	+

• 思考: 能否用更低的计算成本获得更准确的表格结构?



从图的角度理解表格

- 表格顺序无关的特点
 - 行、列位置可交换
 - 更类似于图(网络)结构

实体	数值属性 A	数值属性 B	文本属性 C
甲	100	7	文本1
Z	80	6	文本2
丙	95	8	文本3

文本1 文本3 80 甲 文本2

丙

研究GNN应用于表格任务的可能性: 表格问答,表格补全,数据检查

次要计划

- 时间规划问题
 - A周二到周四上午不忙, B只有周三下午有要事, C是必须参会的领导...
- 设计可以为团体成员安排日程的系统
 - 兼顾事件重要程度
 - 兼顾主要人员参与情况
 - -给出多种备选方案
 - 出现突发事件可以合理调整所有人的日程
- 通过改良时空数据库课程学到的算法实现