

2020 ~ 2021 第一学期汇报

汇报人：赵前

Outline

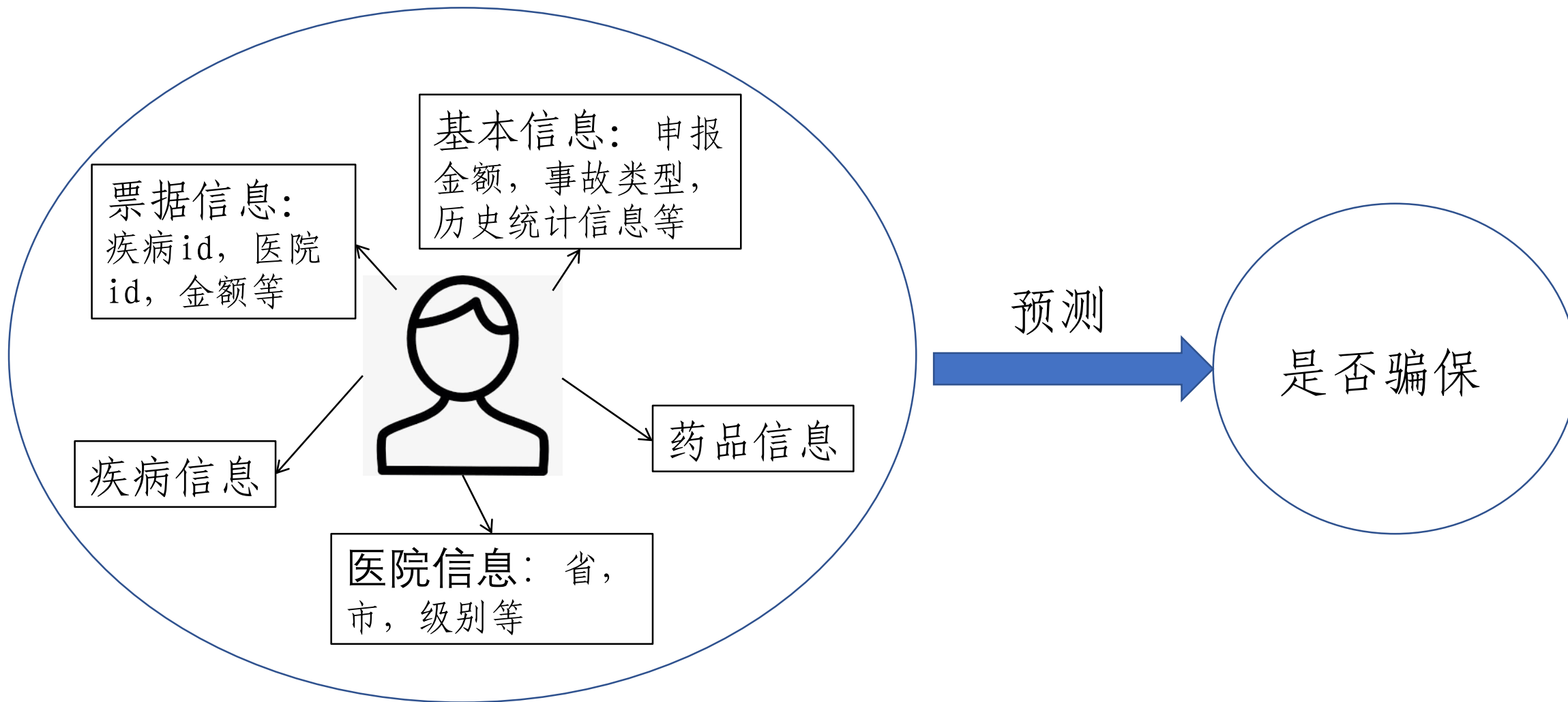
➤ 比赛内容

➤ 科研内容

➤ 项目内容

➤ 未来计划

比赛内容 - 医保欺诈风险预测（众安黑客松）

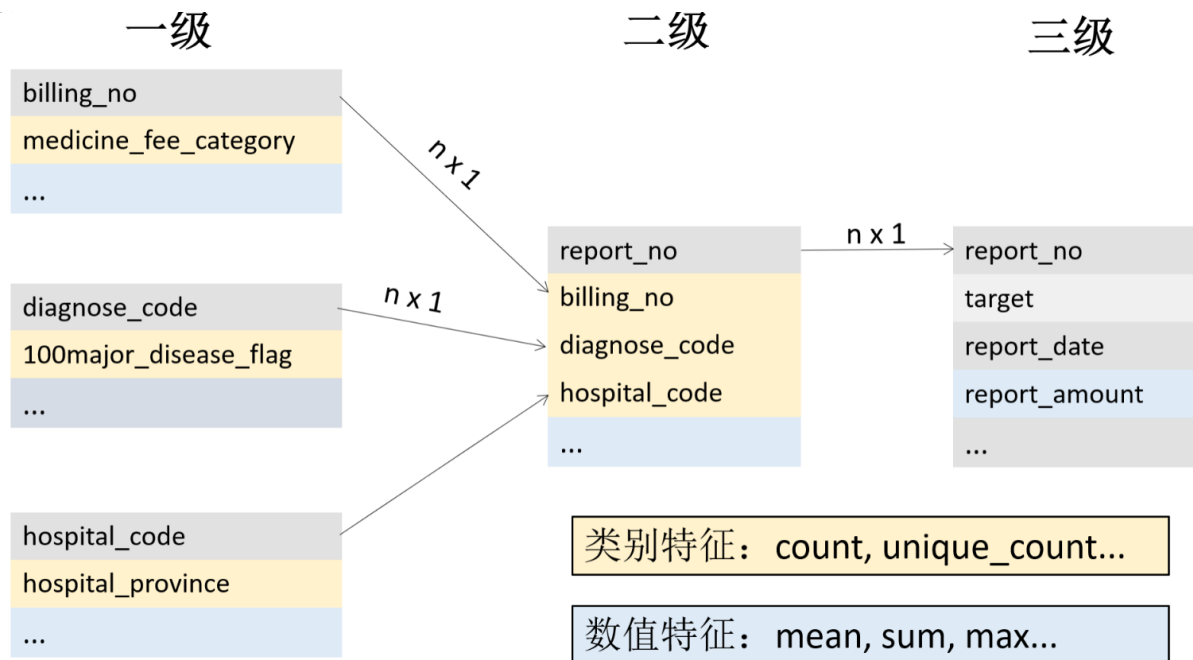


比赛内容 - 成果

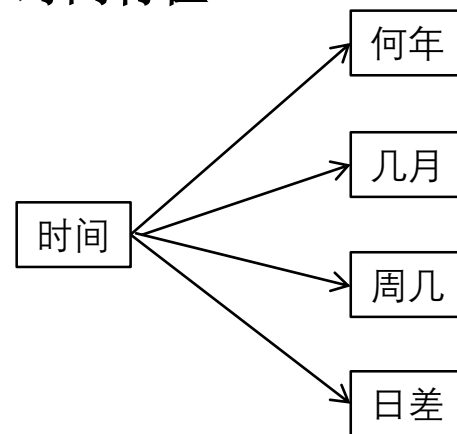
基于企业真实数据，联合时间特征、级联特征等，经自编码器去噪后用决策树预测答案。

最终取得了初赛亚军🏆，决赛第四的成绩

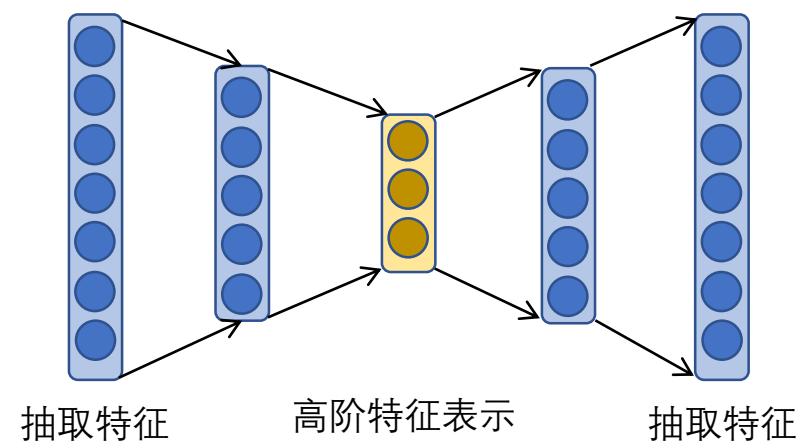
联级特征抽取



时间特征



自编码



比赛内容 - 挑战

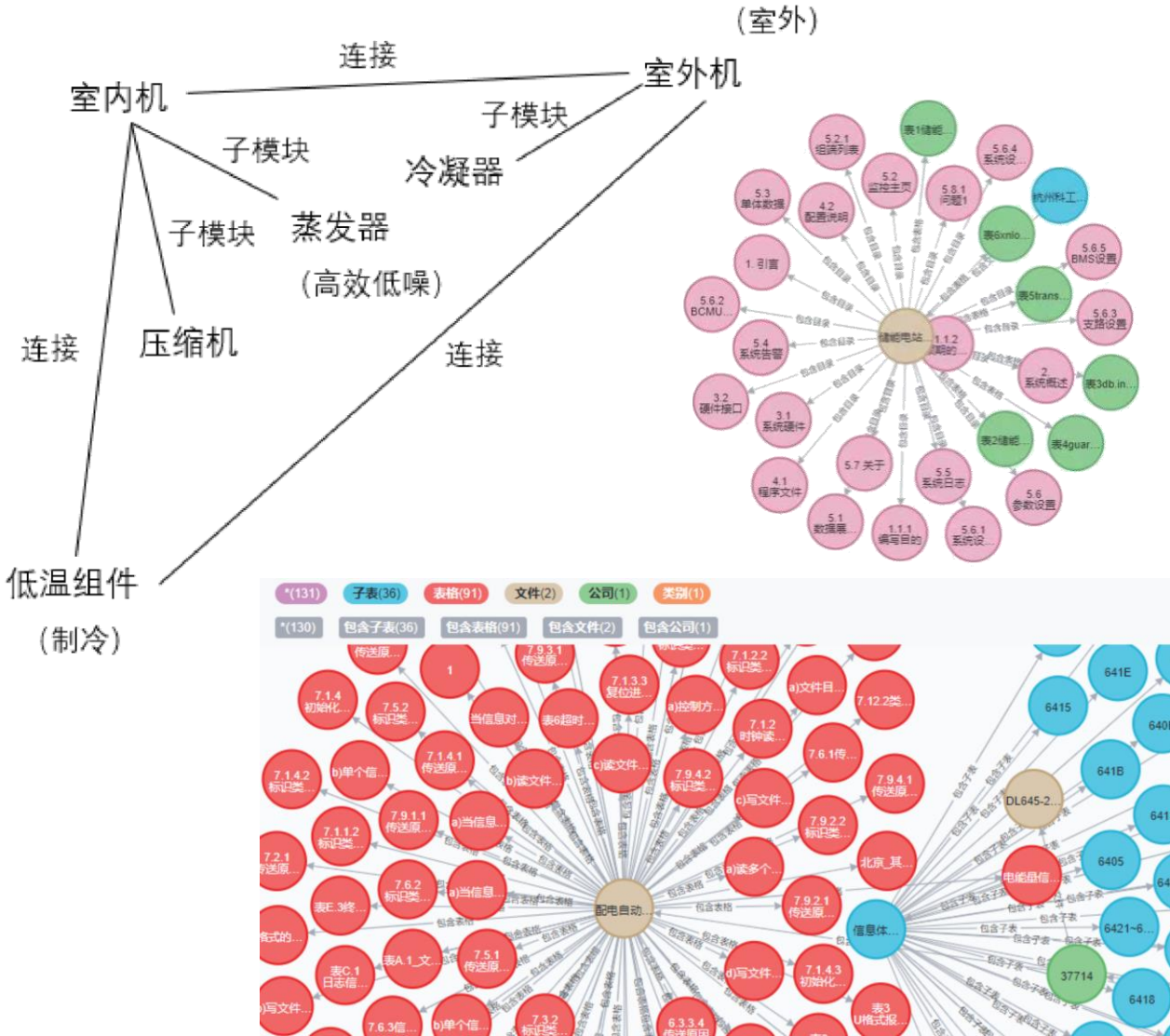
- 原始数据清洗（日期表示方法、年龄区段化等）& 缺失值处理
- 根据特征重要性，探寻有益的特征组合，并利用自编码器进行特征降维以及噪声过滤
- 正负样本不均衡，对深度学习干扰很大，设计采样规则及模型集成等缓解干扰

科研内容

- 探索了适合多轮对话的图模型结构，并构建模型相比 **baseline** 有了些许提升
- 探索将外部知识注入文本的方法，且使用分层过滤的思想降低噪声
- 以物理游戏引擎为基础，探索从模拟环境中对齐知识与状态轨迹之间的关系

项目内容 - 基于知识图谱技术的数据资产管理系统

- 面向工业界非结构化设备文档
- 从零抽取信息并生成粗细粒度图谱
- 基于此完成智能问答系统



项目内容 - 挑战

- 分析富含知识的来源 & 如何构建图谱可提高下游问答的召回率（细粒度表格、粗粒度目录）
- 针对 PDF 中的表格进行识别、跨页表合并、表名称检测等，最终如何转化为细粒度图谱
- 针对问题，如何分词 + 相似度匹配找出图中候选节点，如何基于候选节点推理决策最终答案

未来计划

科研计划:

- 探索反常识问句（如：太阳有几只眼睛？）的鉴别
- 完成毕业设计相关的实验与论文

项目计划:

- 完成项目中的收尾工作，根据甲方意见完善系统
- 完成项目所需论文及专利

THANKS