

科研团队量化

1 数据初步分析

- 项目名称关键词提取/李其炎
 - 停用词去除 — 去除与主题关联性不高的停用词，如“方法”“应用”等等
 - 专用词汇添加 — 添加分词过程中淹没的专业词汇，如“大数据”“区块链”等等
 - jieba关键词的提取库
 - 语句分词
 - 关键词提取

- 知网学科爬虫/陈坤
 - 发送请求
 - 目标站点：知网全部检索结果分析
 - 请求参数：项目名称关键词
 - 获取检索结果 — 获取学科分布内容
 - 解析响应内容
 - 解析排名前十学科
 - 解析学科对应占比及文献数量
 - 保存所解析数据结果

- 承办学院程度匹配/李月
 - 获取每个项目承办单位
 - 根据承办单位进行相应的学科专业获取
 - 学校官网寻求承办单位信息并进行相关学科提取
 - 过程注意干扰词汇（如‘学院’、‘与’）去除
 - 进行知网学科和承办单位学科的相似文本匹配
 - 对每个项目知网解析的所有学科均进行匹配
 - 提取最高匹配度作为匹配率

- 项目学科综合匹配程度/李月
 - 获取匹配到的学科数目并分情况进行处理
 - 一个学科情况 — 该学科即为项目所对应的学科
 - 两个学科情况 — 由两个学科（学科1，学科2）的知网和承办单位匹配率计算得出 — 知网匹配率*0.4+承办单位匹配率*0.6
 - 学科1与学科2计算结果相近且都接近0.5：两个学科都选取
 - 学科1与学科2计算结果差异较大：选取计算结果较大所对应的学科
 - 三个及以上学科情况 — 按照知网匹配率和承办单位匹配率由高到低分别对学科进行排序，得到知网排序和承办单位排序 — 取两个排序结果中前三个学科的交集m
 - 交集个数大于等于2 — 交集部分的学科即为项目对应的学科
 - 交集个数小于2 — 选取知网排序前三学科、承办单位排序前三学科 — 知网前三学科中选取其承办单位匹配率最高的学科x 承办单位前三学科中选取知网匹配最高的学科y — 最终学科：学科x,学科y,交集m取并集并去重

2 量化建模与求解

- 评价指标的建立/李月、李其炎
 - 年均科研经费（0.5）
 - “总科研经费/项目进行时间”获得单项目年均经费 — 外协部分项目进展时间近似1年处理
 - 多项目的年均经费取平均获得团队单项目的年均科研经费 — 项目进展时间不足一年按照一年处理
 - 团队项目数量（0.3）
 - 外协、横向、纵向各团队项目数量
 - 团队项目总数量
 - 团队跨学科数量（0.2） — 以项目为单位提取每个项目对应学科
 - 以负责人工号为参数调用groupby方法，进行分组聚合
 - 横、纵、外协分组聚合后进行pd.concat拼接 — 拼接后运用list({}.fromkeys().keys())去除多项目合并重复学科
- TOPSIS模型建立/陈坤、李其炎
 - 数据预处理 — 对三指标数据进行向量规范化
 - 加权向量规范化 — 初步设定权重向量为 $w = [0.2, 0.5, 0.3]$
 - 计算正理想解与负理想解
 - 计算各项目到达各理想解的距离
 - 建立模型指数对不同项目进行排序

3 模型应用

- 项目数据预处理/陈坤、李月、李其炎
 - 学科映射 — 建立学科映射表，将知网学科划分映射至标准学科划分
 - 分学科统计 — 以不同学科为标准，将各项目划分至各学科下分别进行统计
- 评价体系应用/陈坤、李月、李其炎
 - 评价指标建立
 - 团队综合评价指数（0.2） — 量化建模与求解获得的团队综合评价指数
 - 各学科下项目数量（0.3） — 分学科统计，本团队涉及到本学科的项目数量
 - 年均科研经费（0.5） — 分学科统计，本团队涉及到本学科的所有项目的年均科研经费
 - TOPSIS模型应用
 - 数据预处理
 - 加权向量规范化
 - 计算正理想解与负理想解
 - 计算各项目到达各理想解的距离
 - 建立模型指数对不同项目进行排序