

大学生创新创业训练计划项目

**创新训练项目中期（阶段）暨展前检查表**

项目编号： 202203005

项目类型：创新训练R 创业训练□ 创业实践□

项目名称: 跨学科合作网络发现与推荐

项目名称（英文）： Interdisciplinary Collaboration Network Discovery and Recommendations

项目依托学院： 计算机学院（国家示范性软件学院）

项目负责人： 胡逸同

联系电话： 13661192968

E-mail： wiederholung@outlook.com

指导教师： 王祎

E-mail： yiwang@bupt.edu.cn

起止年月： 05/2022-05/2023

填报时间： 2023 年 2 月 27 日

**填写说明**

1. 本检查表所列各项内容均须实事求是，认真填写，表达明确严谨，简明扼要。
2. 检查表为大16开本（A4），根据内容需要可自行加页，但格式须与原件一致。
3. 检查表填写完毕后，须在“北京邮电大学大学生创新创业计划训练平台” 端口开放时间段内上传并提交，并请项目指导教师在网站上进行审核。
4. 提交检查表前，要确定所有团队成员和指导教师已经在“北京邮电大学大学生创新创业计划训练平台”网站上注册完成，否则无法正常提交。
5. 检查表填写时，“项目组成员签字”“指导教师签字”“学院检查小组意见”“学校意见”无须填写。

**北京邮电大学大学生创新训练项目中期（阶段）暨展前检查表**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 项目名称 | 跨学科合作网络发现与推荐 | | | | | | |
| 项目负责人 | 姓名 | | 联系电话 | | 专业 | 学号 | 邮箱 |
| 胡逸同 | | 13661192968 | | 电子商务及法律 | 2020213350 | wiederholung@outlook.com |
| 项目组成员 | 姓名 | | 联系电话 | | 专业 | 学号 | 邮箱 |
| 王伊哲 | | 15353519783 | | 电子商务及法律 | 2020213355 | anonymity@bupt.edu.cn |
| 巢环宇 | | 15580092683 | | 电子商务及法律 | 2020213349 | chaohuanyu@bupt.edu.cn |
| 朱子炫 | | 13935559636 | | 电子商务及法律 | 2020213363 | ternura@bupt.edu.cn |
| 王俊翔 | | 18910761281 | | 电子商务及法律 | 2020213354 | 1034696975@bupt.edu.cn |
| 指导教师 | 姓名 | | 职称 | | 学院 | 电话 | 邮箱 |
| 王祎 | | 教授 | | 计算机学院（国家示范性软件学院） | 13127958790 | yiwang@bupt.edu.cn |
| 张高唯 | | 讲师 | | 计算机学院（国家示范性软件学院） | 18817872862 | zhanggaowei@bupt.edu.cn |
| 项目级别  （定级阶段） | A级 □ B级 □ C级 R D级 □ | | | | | | |
| 项目实施时间 | 立项时间： 2022年 5 月 预计结题时间： 2023 年 5 月 | | | | | | |
| 1. 项目进展情况：   项目首先对于上一阶段未完成的**项目功能进行添加**，明确了需求并优化改进了系统功能：  图（1）  按照项目规划，我们在上一阶段构建学术合作数据库，并在数据库中分析和梳理其中的跨学科学术合作关系网络的基础上，使用PCA老师处理的**特征**，然后根据其特征进一步进行分析和学习推理，给出用户潜在的科研学术合作伙伴推荐。我们按照以下架构图，对我们整体系统进行了完善工作：  图（2）  1. 推荐算法的设计  首先我们基于上一阶段已经完成的关于老师的**数据库，**在多个高水平期刊/会议（ACM, IEEE, CVPR, Springer等）上对于老师的出版物进行检索并爬取所有论文有关的**摘要**，通过词嵌入的方式进行词向量提取，将其作为老师的特征值。之后使用特征值对于这些老师进行**聚类**，将具有隐合作关系的老师们作为一个簇。在老师搜索相关推荐时，同一类簇会显示在合作推荐图上进行展示。  2. 数据可视化设计与技术实现  图（3）    本项目前端设计主要以 VUE 为载体，并使用 Echarts 表格进行辅助呈现。  **本项目已经实现了六张数据透视表及相关功能**。首先对于中间的学者合作关系网图，此图展示了以学者12为核心的**合作关系网**。用户可以通过将鼠标放置于点上来查询学者的基本信息或者放置于点间连线上来查询学者间的合作关系的详细信息。  左侧展示栏包含三张表格。从上到下依次为学者主要研究方向及合作推荐表、学者个人发表趋势表、学者总体论文数量表。  对于学者主要研究方向及合作推荐表，本项目使用一张旭日图来向用户展示目标学者在大领域与细分研究领域上的研究方向。通过此图用户可以直观了解到目标学者在各个大领域与细分领域上的研究成果比例。此图最外层向用户展示了在各个细分领域下针对目标学者的潜在合作者推荐，学者可以依据此图寻找并联系潜在合作者。  对于学者个人发表趋势表与学者总体论文数量表，本项目使用折线图来向用户展示目标学者论文年发表数量与总论文发表数量的变动情况。通过学者个人发表趋势表用户可以直观了解到目标学者年论文发表数量及论文发表数量的变化趋势。通过学者总体论文数量表用户可以直观了解到目标学者总体论文发表数量及论文总体发表数量变化趋势。  右侧展示栏包含一张表格。其为为相关论文表。本项目提取了目标学者影响力较大的几篇文章向用户展示。用户可以依据此表初步了解该学者的研究方向、研究领域与学术成果。  3. 部署在服务器  将设计好的网站更新在我们部署的服务器上面，通过自购域名可以**直接进行访问**。后端选用 Flask 最为服务器，响应 URL 并返回前端所需的 json。期间解决了开发环境下“非本机访问”和“ CORS 跨域访问”。  4. 论文与专利工作  我们通过分工，并行地展开了发明专利与会议论文的撰写工作，详见下文。 | | | | | | | |
| 二、研究成果概述（含发表文章或申请专利）（300字以上）  1. 数据获取  我们根据北邮九大学院的特点，分别选择了认可度高的**期刊/会议**，编写了相应的摘要爬虫代码，并开源在项目仓库上。  2. 特征工程  我们通过GloVe模型对学者的所有摘要进行词嵌入处理，得到一个50维的特征向量作为学者的表征，便于后续的分析与处理；另一方面，我们参照中科院分区表，对各学者的研究领域进行提取分类。  3. 发现与推荐算法  我们对50维的向量进行了主成分分析（PCA）以降低向量维度，然后进行了聚类（DBSCAN）。下图是计算机学院老师的聚类结果，该聚类结果表征学者间的**隐合作关系**，可视化在合作推荐图上。  图（4）  4.可视化呈现  我们根据聚类结果对可视化界面进行美化，分别对学者的数据进行呈现，详见**图（3）**  5. 论文与专利工作  **专利**：本系统在跨学科合作网络发现与推荐的可视化表达和算法两部分做出创新性工作，固计划申请两篇发明专利。目前正在着重修改网络可视化专利。  **论文**：同样的，我们也计划发表两篇会议论文，一篇介绍我们提出的面向学者的**图嵌入方法与分析方法**（暂定SMC 2023），另一篇介绍我们将**合作网络可视化表达**的方法（暂定ICFEICT）。 | | | | | | | |
| 1. 下一阶段工作计划：   在后续的设计改进过程中，我们将做到以下几点：第一，在教务处的帮助下，**与愿意参与测试的老师们取得联系并检测使用效果**，为之后的项目推广做好铺垫，同时照顾到部分老师不愿意透露个人信息等个人情况，来针对不同老师进行进一步优化；第二，**继续完善推荐算法**，改进算法中对于特殊情况的疏漏并进一步**提升推荐速度**；第三，进一步完善我们的可视化界面，使得页面上的交互更加人性化，页面设计更加美观。 | | | | | | | |
| 1. 存在的问题、建议及需要说明的情况：   本项目可能存在隐私问题，需要调查学校需求和老师评价后进一步调整。  拟计划引入分布式的学习方法（联邦学习），确保各学校教务的教师数据不出本地，在本地训练模型并将权重和梯度加密后聚合到云端，得到联合模型。 | | | | | | | |
| 1. 指导教师综合评价：   本项目进展顺利，在项目开展的过程中，项目组同学能够积极学习新的技术知识体系，提出解决问题的创新思路，在项目执行的过程中，展示出较强的执行力，针对项目中出现的状况，能够迅速响应和解决，技术能力和创新能力都得到了极大的提升。目前的系统原型较好的实现了预期目标，能够保证预期成果的完美实现，建议继续努力，实现项目的完美交付和项目的结题。 | | | | | | | |
| **项目组成员个人分工及完成情况** | | | | | | | |
| 姓名 | | 完成情况 | | 工作内容 | | 工作量 | |
| 胡逸同 | | 已经完成 | | 聚类算法拟合，对于数据进行标签和实现 | | 每周7h开发 | |
| 王伊哲 | | 已经完成 | | 编写可视化算法与实现 | | 每周7h开发 | |
| 朱子炫 | | 已经完成 | | 页面架构与设计 | | 每周7h开发 | |
| 王俊翔 | | 已经完成 | | 完成最终前端页面 | | 每周7h开发 | |
| 巢环宇 | | 已经完成 | | 专利撰写 | | 每周7h开发 | |
| **经费使用情况：**  **阿里云OSS：30/年**  **阿里云ECS 285.18/1年**  **域名metattri.com 61/1年**  **VPN 1 170/年**  **VPN 2 200/年**  **合计：746.18￥** | | | | | | | |
| 项目组成员签字： | | | | | | | |
| 指导教师签字：  2023年 2月 27日 | | | | | | | |
| 学院检查小组意见：  基地主任签字（章）： 年 月 日 | | | | | | | |
| 是否同意继续执行 | | | | 是 否 | | | |
| 学校意见：  签章： 年 月 日 | | | | | | | |