“智慧税务”小组夏令营科研实训题目

**2024/06/19**

* 科研实训题目

请同学在如下范围内任选一个题目。

1. 知识图谱增强的的大模型训练（提升大模型推理、行动能力等）
2. 大模型的数学能力增强
3. 从大模型的隐式知识中蒸馏出知识图谱（例如商品图谱，包含商品上下游关系，品类关系，替代关系、互补关系等）
4. 大模型的反学习（如何忘掉一个东西）
5. 标签噪声对可解释性算法的影响及解决方案探究
6. Graph 深度学习的不确定性度量
7. Graph 图分类任务的OOD检测问题
8. 基于大模型的多模态数据增强（如何利用大模型技术自动增强出高质量标注数据，如何挑选最有价值训练的训练数据进行增强降低生成成本与训练成本）
9. 基于大模型的少样本学习方法（如何利用大模型能力从领域海量无标注数据中选择少部分最优价值的难样本进行标注和微调，降低标注成本和训练效率）
10. 多模态任务中的图文对齐方法（在缺乏人工标注信息情况下，如何实现文本中实体与图像中object的精准对齐，提升下游多模态分类、抽取等任务性能）
11. 基于大模型的自举式思维链微调（在无人工或低成本标注开销下，利用大模型自身能力对只有X,Y标注的训练集增强出X到Y的思维链，利用增强后的思维链数据微调模型提升性能）
12. 图数据与大语言模型的结合，图基础模型（Graph Foundation Models GFM）的构建与下游应用
13. 其他大模型和机器学习鲁棒性（让机器知道自己什么时候不知道）的研究，主题相关题目自拟也可以

* 考核形式

1. 调研自选题目相关的代表性论文，要求为CCF A类论文；
2. 论文数量不限，最好能从调研中得到这个小方向的脉络和一些思考；
3. 要求对其中至少一篇最新工作进行复现；
4. 欢迎大家在调研中与老师交流，看看读研后老师是如何指导大家科研的；
5. 独立进行10min（7min 汇报+3min 提问）的PPT 考核在线答辩；
6. 所有材料（答辩PPT 与工程代码等）压缩打包提交，命名为“姓名\_学员考核.zip”。

* 考核时间

与团队其他小组考核时间相同，请注意通知。

* 小组导师

郑庆华、师斌、彭祯、孙凯

联系方式：微信群

