OpenKS 2020年度工作计划 Roadmap

一、任务制定原则

采取应用驱动方式，以实现基于领域知识图谱的问答与推荐功能为目标，解构主流图谱问答（KGQA）和图谱推荐（KGRec）流程各环节，结合已形成的知识计算主要功能模块设计和内部已有技术工作基础，形成完整的2020年度工作计划Roadmap，最终实现OpenKS-Alpha版本（v0.1.1），打通从数据到知识到应用的完整链路。

二、系统应用清单

* 知识图谱问答（KGQA）：面向2个以上场景领域数据，支持1种以上典型或创新知识图谱问答流程的设计和开发，整合封装内部可支持分布式的算法模型，发布可供二次开发集成的知识图谱问答接口，输入为文本问题，输出为多模态知识答案。
* 知识图谱推荐（KGRec）：面向2个以上场景领域数据，支持1种以上典型或创新知识图谱推荐流程的设计和开发，整合封装内部可支持分布式的算法模型，发布可供二次开发集成的知识图谱推荐接口，输入为某实体，输出为针对该实体被推荐的实体集合。

三、2020 Roadmap

| 时间范围 | 工作内容 | 提交物 |
| --- | --- | --- |
| 6月 | 完成Paddle分布式代码的阅读 | 代码分析报告提交，包括部分C++代码 |
| 6月 | 完成项目模块基本搭建与基础类开发，包括问答/推荐应用+模型调用+模型开发 | 模型集成与调用规范设计文档及相关代码 |
| 6-7月 | 完成某领域数据集准备、标注、预处理开发，包括问答语料/图谱数据/多模态数据 | 数据抽象及处理模块代码集成到项目中 |
| 6-7月 | 完成分布式GPU集群和平台实验环境的搭建 | 可基于Paddle运行分布式GPU模型训练的环境 |
| 7月 | 完成分布式机器学习在领域图谱构建与推理场景中的使用调研 | 主流分布式算法测试报告，调研分析报告等 |
| 7-8月 | 完成基于知识图谱的问答和推荐应用流程调研和初步pipeline设计 | 调研分析报告，代码基本框架搭建和集成 |
| 7-10月 | 基于某领域多模态数据和Paddle框架的知识图谱构建，相关算法模型研发 | 算法模型代码集成到对应项目模块中 |
| 7-10月 | 基于某领域知识图谱数据和Paddle框架的图表示学习推理算法模型研发 | 算法模型代码集成到对应项目模块中 |
| 8-9月 | 完成项目中分布式模块基本框架搭建与基础类开发，包括底层通信架构接口+上层算法优化接口 | 主流分布式机器学习算法与接口集成到项目中 |
| 9-10月 | 完成基于已有开源分布式机器学习算法对内部已集成自主研发算法模型的训练测试 | 分布式信息抽取+分布式图谱表示的测试报告 |
| 10-11月 | 完成基于自主研发算法模型的问答和推荐demo接口组装开发 | 模型调用+问答推荐特有模型开发+整体pipeline封装代码集成 |
| 11-12月 | 完成知识计算场景下分布式模型训练算法优化的方案设计 | 制定下一年重点攻关研究任务，做相关汇报和评审 |
| 12月 | 完成OpenKS的完整可运行知识计算流程封装和使用 | 可利用自主研发图谱构建和图表示学习算法模型进行领域问答和推荐演示 |
| 12月 | 发布OpenKS v0.1.1版本，即OpenKS-Alpha | OpenKS-Alpha项目代码及相关文档 |

四、项目开源地址

组织名：ZJU-OpenKS

项目仓库名：OpenKS

地址：<https://github.com/ZJU-OpenKS/OpenKS>