

基于内容的科技文献推荐算法的研究

一、内容概述：

本论文研究基于内容的科技文献推荐算法，通过对科技文献内容的分析，深化科技文献的特征表达，从而提高推荐系统的准确度。

二、选题的目的及意义：

基于内容的科技文献推荐算法，在使用户快捷准确地获得所需学术资源方面拥有巨大潜能。本课题能够从海量信息中向用户自动推荐出符合科研需求的资源，这种技术使得网站从以信息为中心到以用户为中心的模式转移，节省宝贵的科研时间。并且本研究提升了推荐算法的精确度。

三、工作内容：

1. 数据采集：按关键字下载科技论文，对论文进行数据预处理，获得文本格式的论文数据集。
2. 特征提取：从数据集中提取题目、摘要和参考文献作为内容特征，对特征进行分词，去停用词处理并向量化。
3. 建立模型：建立基于内容的摘要，题目，引用关系模型。
4. 设计和实现算法：构造论文向量，利用相似度分类或推荐，并通过对比试验验证结果。
5. 根据研究结果，撰写毕业论文。

四、进展与创新：

1. 改进了基于内容的文献推荐算法。即在该算法基础上融合基于引用关系的推荐算法，为用户推荐相关领域内最新的研究方向。
2. 使用深度学习针对论文内容特征提取任务构建了特征工程。深度学习因其强大的学习数据集本质特征的能力能够获取用户和项目的深层次特征表示，且能够获得数据的统一表征。将深度学习与传统推荐方法结合进行科技文献推荐，能够有效利用多源异构数据，缓解传统推荐系统中的数据稀疏和冷启动问题。