**公示情况表**

**1.项目名称**：

网络大数据敏捷计算与精准服务关键技术及产业化应用

**2.推荐单位（专家）：**

华南师范大学

**3.项目简介：**

本项目属于数据库技术领域。

在“十三五”规划中，大数据已上升为国家战略，它正对各行业产生重大影响，而网络大数据是其重要组成部分。本项目针对不同组织在智能化信息化过程中，涉及的网络大数据管理、计算与服务等方面的棘手问题，重点依托863计划、国家自然科学基金等项目，以先进的知识图谱和深度学习理论方法体系为驱动，通过产学研协同攻关和全方位的集成示范应用，形成了涵盖网络大数据规范组织与知识管理、敏捷近似计算、精准信息推荐以及深层主题发现与趋势预测等在内具有自主知识产权的创新性成果，突破了网络大数据从来源、管理到应用全生命周期所面临的技术瓶颈。

具体来说，项目主要创新成果如下：

1．网络大数据多模态个性化信息推荐：针对现有推荐系统存在冷启动、泛化能力差和个性化不足等难题，首次揭示了文本和图片联合空间下大规模物品的分布式表征机理，进而发明了融合Attention机制和深层神经网络的多模态信息推荐模型，并最终研发了基于深度强化学习的推荐过程建模技术来最大化信息推荐的全局收益。

2．网络大数据近似计算与敏捷分析：突破了用户使用SQL/HiveQL困难的技术瓶颈，实现基于自然语言接口的交互式查询分析方法，进而，首次发明了计算精度与时间代价均衡的网络大数据聚集分析和多目标查询技术。

3．网络大数据组织、存储与知识管理：针对不同组织机构中非结构化网络大数据存储效率低下难题，提出了三层压缩NSB树的极小路径存储方法来显著降低空间开销，同时，集成研发了4种知识图谱优化技术来解决其构建过程中存在的可信度和质量问题，并首次发明了基于知识图谱的词嵌入深度学习方法来大幅度提高词语分布式表征的语义准确度。

4．网络大数据深层主题发现与趋势预测：独创一种面向不同应用的深度信息识别框架，攻克了不同应用下网络大数据信息识别模型不统一的难题，同时，首次发明了基于知识图谱的深层主题检测技术来解决现有方法在捕获数据语义信息和自动确定主题数量等方面存在的缺陷，在此基础上，实现了融合知识图谱和模糊隐马尔可夫模型的主题演化趋势精准预测方法。

项目已获授权发明专利38项、实用新型专利12项，申请发明专利31项，获软件著作权66项，主编和参编国家/地方技术标准规范12部，发表学术论文243篇，SCI/EI检索192篇（其中ACM/IEEE Trans.论文36篇、CCF-A类会议论文12篇），google学术总他引3100余次，7篇论文先后获得国际会议优秀论文奖，作国际会议特邀报告11次，出版专著8部，成果得到世界30余个国家和地区学者的引用和正面评价。

经教育部科技成果鉴定认为项目成果总体达到国际先进，部分居国际领先水平。

项目成果在上海、北京、广东、江苏等全国近30个省、市自治区的国有大中型企业、机关事业单位与民营中小型企业的4000余项工程项目中得到成功应用。近三年，项目新增产值14.82亿元、新增利润3.65亿元，并创造良好的社会效益。同时，项目成果在国家大数据和人工智能战略背景下，具有广阔的应用前景。

**4.主要完成单位及创新推广贡献、推广应用情况：**

**1）主要完成单位：**

华南师范大学、同济大学、上海师范大学、中国矿业大学、上海中信信息发展股份有限公司、安徽博约信息科技股份有限公司、神华和利时信息技术有限公司

**2）创新推广贡献、推广应用情况：**

项目集成研发了网络大数据敏捷计算与精准服务平台，兼顾其产业化应用。该项目成果已先后成功应用于上海、北京、广东、江苏、浙江、山东、福建、湖北等全国近30个省、市自治区的4000余项工程项目中，服务对象包括国家发改委、最高人民检察院、最高人民法院、中国移动通信集团及其所属子（分）公司、中国化工集团公司、国家能源集团（原神华集团部分）及其二级子公司、国家档案局、上海市公安局等在内的国有大中型企业、机关事业单位以及中小型企业。近三年，本项目累计新增产值14.82亿元、新增利润3.65亿元、新增税收0.53亿元、节支总额1.04亿元，推动我国多家组织机构实现了智能化信息化科学管理的快速转变，优化了业务流程，提高了网络大数据存储、计算和服务的效率，产生了巨大的经济和社会效益。

近年来，随着云计算、移动通信和社交网络等技术的迅猛发展，我们所能获得的数据正以超出想象的速度快速增长。根据国际数据公司IDC（International Data Corporation）的测算，目前全球数据的增长速度在每年40%左右，到2020年，数据总量将达到40ZB。当前，在国家大数据战略和国家人工智能战略的背景下，我国各行各业都在持续加大不同网络来源大数据的存储、管理、计算与服务的研发投入和专业人才培养，旨在加快推动数据资源共享开放和开发应用，助力产业转型升级和社会治理创新。根据我们调研，金融、电信、医疗、教育、能源等5大行业以及政府机关在2018年的大数据研发和应用投入就超过了4000亿人民币。而本项目的核心技术及创新性成果突破了网络大数据从管理、计算到服务全生命周期所面临的技术瓶颈，正好迎合了不同行业大数据研发和应用的技术需求，具有广阔的推广应用前景。

**5.曾获科技奖励情况：**

无。

**6.主要知识产权证明目录：**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **知识产权类别** | **知识产权具体名称** | **国家**  **（地区）** | **授权或申请号** | **权利人** |
| 授权发明专利 | 一种基于用户网络数据的物品信息推荐方法 | 中国 | ZL201410403849.3 | 同济大学 |
| 授权发明专利 | 一种社交网络的物品评分及推荐方法 | 中国 | ZL201510228294.8 | 同济大学 |
| 授权发明专利 | 一种社交网络环境中的用户间推荐信任度计算方法 | 中国 | ZL201210442457.9 | 上海师范大学 |
| 授权发明专利 | 基于层次选择性视觉注意力机制的 JND 阈值计算方法 | 中国 | ZL201310563526.6 | 同济大学 |
| 授权发明专利 | 文件格式转换方法及转换工具 | 中国 | ZL201110021467.0 | 上海中信信息发展股份有限公司 |
| 授权发明专利 | 一种务联网的场景感知时态信息管理方法 | 中国 | ZL201110403676.1 | 华南师范大学 |
| 授权发明专利 | 基于决策树规则和多种统计模型相结合的人名识别算法 | 中国 | ZL201410060957.5 | 安微博约信息科技股份有限公司 |
| 授权发明专利 | 一种网络水军的探测与判定方法 | 中国 | ZL201210050176.9 | 安微博约信息科技股份有限公司 |
| 授权发明专利 | 一种基于拓扑势的重叠复杂网络社区发现方法及系统 | 中国 | ZL201310392149.4 | 中国矿业大学 |
| 授权发明专利 | 面向文本大数据的词语处理方法 | 中国 | ZL201410247336.8 | 神华和利时信息技术有限公司 |

**7.主要完成人情况表：**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **姓名** | **排名** | **技术职称** | **工作单位** | **完成单位** | **对本项目技术创造性贡献** | **曾获科技奖励情况** |
| 黄震华 | 1 | 教授 | 华南师范大学 | 同济大学 | 创新点（一）、（二）、（三）、（四）。 | 无 |
| 汤庸 | 2 | 教授 | 华南师范大学 | 华南师范大学 | 创新点（二）、（三）。 | 无 |
| 张冬冬 | 3 | 教授 | 同济大学 | 同济大学 | 创新点（一）。 | 无 |
| 张波 | 4 | 教授 | 上海师范大学 | 上海师范大学 | 创新点（一）、（四）。 | 无 |
| 王志晓 | 5 | 教授 | 中国矿业大学 | 中国矿业大学 | 创新点（四）。 | 无 |
| 张曙华 | 6 | 教授级高工 | 上海中信信息发展股份有限公司 | 上海中信信息发展股份有限公司 | 创新点（二）、（三）。 | 无 |
| 黄昌勤 | 7 | 教授 | 华南师范大学 | 华南师范大学 | 创新点（一）、（二）。 | 无 |
| 郑中华 | 8 | 教授级高工 | 安微博约信息科技股份有限公司 | 安微博约信息科技股份有限公司 | 创新点（三）、（四）。 | 无 |
| 潘涛 | 9 | 教授级高工 | 神华和利时信息技术有限公司 | 神华和利时信息技术有限公司 | 创新点（四）。 | 无 |
| 张琪 | 10 | 教授级高工 | 神华和利时信息技术有限公司 | 神华和利时信息技术有限公司 | 创新点（四）。 | 无 |
| 杨安荣 | 11 | 教授级高工 | 上海中信信息发展股份有限公司 | 上海中信信息发展股份有限公司 | 创新点（二）、（三）。 | 无 |
| 傅城州 | 12 | 讲师 | 华南师范大学 | 华南师范大学 | 创新点（一）。 | 无 |