**中国人工智能学会-华为**

**MindSpore学术奖励基金项目申请书**

**项 目 名 称**

基于大规模图数据的图神经网络结构搜索研究

|  |  |
| --- | --- |
| **申请者姓名** | 高建良 |
| **所属机构** | 中南大学 |
| **提交日期** | 2020年12月1日 |

|  |  |
| --- | --- |
| **申请人姓名** | 高建良 |
| **申请项目名称** | 基于大规模图数据的图神经网络结构搜索研究 |
| **所属学校 / 科研院所** | 中南大学 |
| **所属院系 / 研究科室** | 计算机系 |
| **研究领域** | 图神经网络、机器学习、行业大数据挖掘 |
| **职称** | 教授 |
| **最高学历及毕业时间** | 博士 |
| **性别** | 男 |
| **出生日期** |  |
| **手机号码** |  |
| **电子邮箱** | gaojianliang@csu.edu.cn |
| **CAAI会员编号** |  |
| **个人简介** | 高建良，博士，中南大学计算机学院教授、博士生导师。先后毕业于国防科技大学和中国科学院大学。近年来主要从事大规模图计算技术及其应用研究，2018年出版**《大规模图数据计算技术》**专著一部。  作为项目负责人，主持了国家自然科学基金项目、中国博士后特别资助项目、计算机体系结构国家重点实验室开放课题等项目。在IEEE Transactions期刊和CCF A/B类会议发表学术论文多篇，并与国内外学者建立良好的合作关系。在大数据处理领域取得一系列成绩，曾任IEEE Conference on Big Data 2016大会主席（General Chair）。作为学术带头人，带领中南大学团队建设了我国第一批数据科学与大数据技术专业，编写出版《Spark大数据编程基础》教材。  **代表性成果列表**：  [1] **高建良**，张丽霞; 大规模图数据计算技术,长沙：中南大学出版社, 2018.（ISBN：9787548733607） (学术专著)  [2] **Jianliang Gao**, Tengfei Lyu, Fan Xiong, Jianxin Wang, Weimao Ke, Zhao Li: MGNN: A Multimodal Graph Neural Network for Predicting the Survival of Cancer Patients. *International ACM Conference on Research and Development in Information Retrieval* (SIGIR),2020.(CCF A类会议)  [3] Xiaoting Ying, Cong Xu, **Jianliang Gao**, Jianxin Wang, Zhao Li: Time-aware Graph Relational Attention Network for Stock Recommendation. *ACM International Conference on Information & Knowledge Management* (CIKM),2020.(CCF B类会议)  [4] Zhao Li, Chenyi Lei, Pengcheng Zou, Donghui Ding, Shichang Hu, Zehong Hu, Shouling Ji, **Jianliang Gao**: Attention with Long-Term Interval-Based Gated Recurrent Units for Modeling Sequential User Behaviors. *International Conference on Database Systems for Advanced Applications* (DASFAA),2020.(CCF B类会议)  [5] **Jianliang Gao**, Ling Tian, Tengfei Lv, Jianxin Wang, Bo Song, Xiaohua Hu, Protein2Vec. Aligning Multiple PPI Networks with Representation Learning, I*EEE/ACM Transactions on Computational Biology and Bioinformatics*, 2019, DOI: 10.1109/TCBB.2019.2937771. (SCI)  [6] **Jianliang Gao**, Jianxin Wang, Jianbiao He, Fengxia Yan: Against Signed Graph Deanonymization Attacks on Social Networks. *International Journal of Parallel Programming*. 47(4): 725-739,2019. (SCI)  [7] Jun Gao, **Jianliang Gao\***, A Similarity Measurement Method Based on Graph Kernel for Disconnected Graphs. IJCAI 2019. (CCF A类会议)  [8] Fan Xiong, **Jianliang Gao\***, Entity Alignment for Cross-lingual Knowledge Graph with Graph Convolutional Networks. IJCAI 2019. (CCF A类会议)  [9] Hongliang Du, Zhiyi Jiang, **Jianliang Gao\***, Who is Who: Name Disambiguation in Large-Scale Scientific Literature. ICDM Workshops 2019: 1037-1044.  [10] **Jianliang Gao**, Jianxin Wang, Ping Zhong, Haodong Wang, On Threshold-Free Error Detection for Industrial Wireless Sensor Networks, *IEEE Transactions on Industrial Informatics*, 14(5): 2199-2209, 2018. (中科院一区SCI，高被引论文)  [11] **Jianliang Gao**, Bo Song, Weimao Ke, Xiaohua Hu, BalanceAli: Multiple PPI Network Alignment With Balanced High Coverage and Consistency,*IEEE Transactions on NanoBioscience*, 16(5): 333-340,2017. (SCI)  [12] **Jianliang Gao**, Bo Song, Zheng Chen, Weimao Ke, Wanying Ding, Xiaohua Hu, Counter Deanonymization Query: H-index Based k-Anonymization Privacy Protection for Social Networks, *International ACM Conference on Research and Development in Information Retrieval* (SIGIR), 2017. (CCF A类会议)  [13] **Jianliang Gao**; Qing Ping; Jianxin Wang, Resisting re-identification mining on social graph data, *World Wide Web*, pp. 1759-1771, 2018. (SCI)  [14] **Jianliang Gao**, Ping Liu, Xuedan. Kang, Lixia Zhang, and Jianxin Wang, PRS: Parallel Relaxation Simulation for Massive Graphs, *Computer Journal*, 2016, 59(6), pp. 848-860.(SCI)  [15] **Jianliang Gao**, Bo Song, Ping Liu, Weimao Ke, Jianxin Wang, and Xiaohua. Hu, Parallel Top-k Subgraph Query in Massive Graphs: Computing from the Perspective of Single Vertex, *IEEE International Conference on Big Data*, pp. 636-645, 2016. |
| **项目概述** | 内容包含但不限于：  一、背景及研究意义（说明项目当前研究背景、研究意义、主要解决的问题）  1.研究背景，  图卷积神经网络-发展脉络  图1图神经网络研究演进历程  图1展示了图神经网络研究演进的部分代表性工作。从2014年早期的谱域图神经网络SCNN到2016年空域图神经网络GCN，图神经网络的计算复杂度得到了有效的改善，2018年的GraphSAGE、GAT进一步提升了图神经网络挖掘图数据的能力，2019年的Cluster-GCN提出一种基于大规模图数据的图神经网络训练方案，2020年提出了GraphNAS图神经网络结构搜索方案，设计了一种集中式的图神经网络结构的搜索方式。  随着图数据的规模增大，大规模图数据上数据特征分布越来越复杂多样，目前没有一种通用的图神经网络结构能在不同的图数据上保证建模的准确性，需要基于具体的图数据针对性地设计图神经网络结构并训练评估模型准确性。然而，人工或穷举法搜索图神经网结构是一个特别耗时的过程。所以，基于图数据特征并行高效地自动搜索出适配图数据特征的图神经网络结构，成为了在大规模图数据上使用图神经网络技术的难点与研究重点。  2.研究意义  图神经网络在不同领域都得到了广泛的应用,例如,电商网络中欺诈用户的检测、自然语言处理中关系抽取、社交网络中链路预测和节点聚类、生物医疗中药物副作的预测，交通预测、程序推断、优化求解等任务上，不用领域产生的大规模图数据具有不同的特征分布，基于不同特征分布的图数据分布式自动搜索图神经网络结构，使用搜索出的结构在不同领域的大规模图数据上建模，能大大提高建模效率与模型准确性，使得图神经网络技术在各个领域的应用更容易实现落地  3.主要解决问题    图2图神经网络结构搜索过程  一个图神经网络结构可以分为多个组成部分。图2给出了一个示例，列出图神经网络每一层（layer）中包含的注意力机制的类型、聚合方式的类型、激活函数等组成部分。以具有6个组成部分的两层图神经网络为例，其结构空间规模的量级为。本项目主要解决的问题，设计分布式图神经网结构搜索算法，基于不同的大规模图数据特征，自适应地在每一个组成部分中高效选择合适的组件形成适应大规模图数据分布复杂多样性的图神经网络结构，达到建模准确性要求。  二、研究目标（根据基金项目的要求，提炼出研究目标）  （1）基础理论创新：本项目将研究基于大规模图数据的图神经网结构搜索研究，力求在基础理论创新上取得突破。重点研究大规模图神经网络结构搜索加速技术，包括基于强化学习与进化学习的并行图神经网络结构搜索算法与基于图划分算法的并行图神经网络训练技术，实现搜索、建模过程的分布式计算，解决基于大规模图数据图神经网络结构搜索加速问题。  （2）应用创新：本项目提出的基础理论创新成果将在实际的大规模图数据中验证,使用MindScope深度学习开发框架将自适应图神经网络搜索功能封装成一个工具库GNAS，结合MindScope提供的分布式建模训练能力，对外提供分布式图神经网络结构自适应搜索能力。  三、研究方法（说明研究项目拟采用的技术路线与研究方法，突出其创新性和可落地性等优势）等。  总体设计方案  图3.总体方案设计  本项目将研究方案将以“初始结构搜索🡪并行搜索加速🡪图应用”为主线，针对基于大规模图数据的图神经网络结构搜索为主要挑战，重点解决在大规模图数据中图神经网络结构搜索加速的关键问题，所有研究与算法开发工作将使用MindScope深度学习框架，具体研究方案如图3所示。  (1).初始结构搜索：  本项目将使用强化学习机制与图神经网络结构相结合，设计符合搜索任务的强化学习智能体与奖励分配机制，将量级为图神经网络结构搜索空间大幅度缩小，为并行化结构搜索提供初始化的小规模图神经网络结构空间。  (2).并行搜索加速：  基于强化学习搜索约束后的图神经网络空间已经得到有效的收缩，为了进一步加速搜索过程，本项目将的利用遗传算法并行搜索的特点，在不同的计算节点上设计不同的遗传算子，基于初始化结构空间分布式并行搜索图神经网络结构，同时使用基于图划分策略的分布式图神经网络训练方案，加速单个图神经网络的训练过程，从而从整体上解决基于规模图数据图神经网结构高效搜索问题。 |
| **论文投稿计划** | 简要描述项目预期产出的论文及投稿计划  (1).在基于大规模图数据的图神经网络结构搜索加速方向取得理论突破，在SIGKDD、IJCAI等重要学术会议发表论文2篇。  (2).将基于MindScope深度学习框架开发自适应图神经网络结构搜索工具包GNAS,并集成到MindScope中。 |
| **主管领导审批意见** | （无需加盖单位公章，确认主管领导意见后注明 “同意”即可） |

注：若有其他需要说明的情况，请以附件形式提供。