（此件为上传附件，字数一般为5000字左右）

一、项目目的意义

1.1研究背景

图（Graph）是一种表示对象之间关联关系的抽象数据结构，包含节点集合和边集合以及它们的属性特征。大规模图数据可以表达丰富的关系。图神经网络可以将不同来源、不同类型的数据融合到一个图里进行分析，从而得到分析孤立数据难以发现的结果。神经网络在很多任务上已经取得了重要的突破，例如图像分类、视频处理、语音识别等，这些任务处理的通常是欧式空间数据。然而在很多应用中涌现了越来越多的非欧式空间数据。为了更好地在分析非欧式数据，图神经网络成为当前的一个研究热点并被广泛应用到工业界[1]。

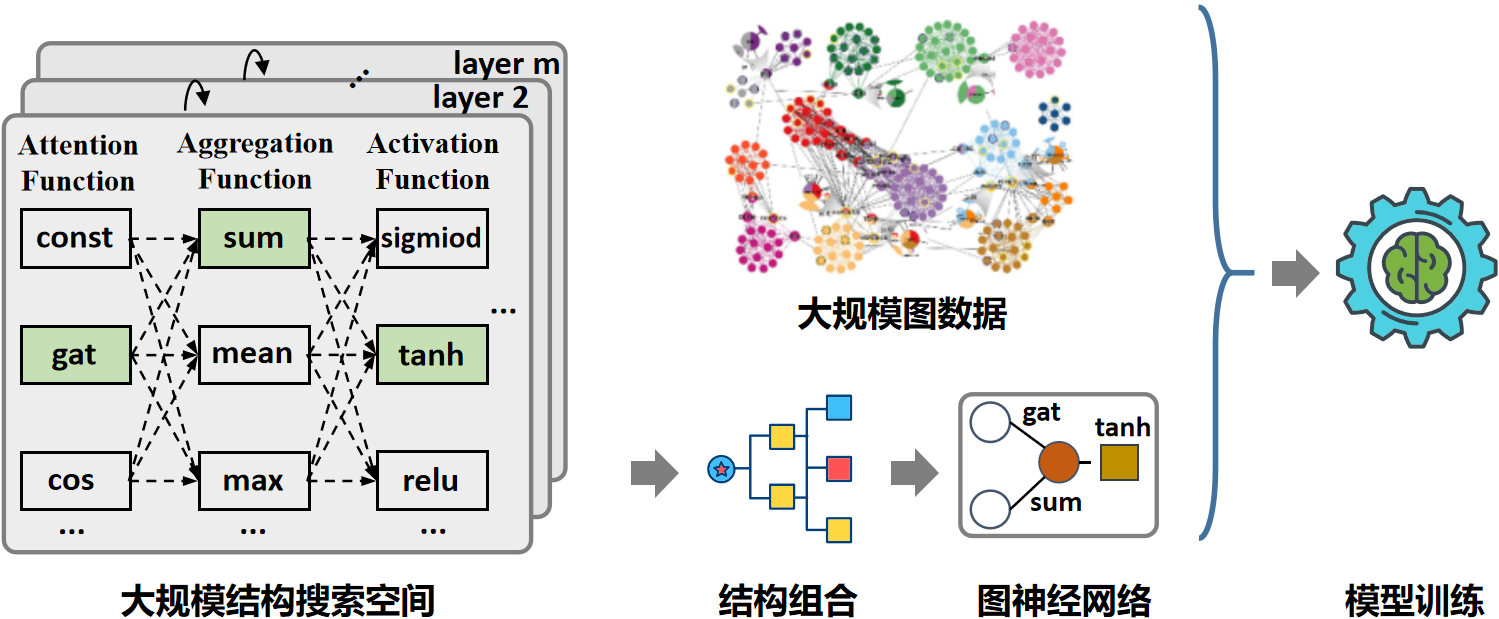


图 1 图神经网络建模过程示例

一个图神经网络由多个结构组件组成[2]。图1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、激活函数等结构组件。图神经网络建模的目标是在每一种结构组件中选择合适的组件形成一个高效的图神经网络结构并在图数据上进行训练获取模型参数的过程。图神经网络建模需要根据不同的图数据特征调整图神经网络结构达到建模准确性的要求。随着实际生产环境中图数据规模与图神经网络结构空间的增加，图神经网络建模面临着复杂度过高计算开销过大的挑战。2019年谷歌提出Cluster-GCN[3]，基于图划分思想对大规模图数据进行分割，为图神经网络在处理大规模图数据上提出了解决方案，但对于不同的图数据集，图划分没有解决建模准确度的问题。由于大规模图数据特征分布复杂度呈上升趋势，需要基于图数据特征构建图神经网络结构。2020年阿里研究团队提出了GraphNAS[4]图神经网络结构搜索方案，设计了一种基于强化学习的图神经网络结构搜索方式，但强化学习存在搜索效率低下无法实现并行等问题，所以无法应用在大规模图数据上。

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，需要基于具体图数据特征针对性地设计图神经网络结构并训练图神经网络评估模型准确性。然而，图神经网络建模是一个特别耗时的过程，以六个结构组件构成的两层图神经网络结构为例，不同图神经网络结构的组合规模在­量级[4]。同时单个图神经网络在大规模图数据上的训练时间与空间开销极大，例如，训练一个GCN[5]网络结构的时间复杂度和空间复杂度分别为和。其中L为图卷积层层数，为邻接矩阵非0元素个数，F为节点特征数，N为节点个数[6]。

高效、自动、并行地从图神经网络结构空间中搜索出适配不同图数据的图神经网络结构，并使用合理的图采样技术保证模型准确性的同时加速单个图神经网络的训练速度，成为了在大规模图数据上加速图神经网络建模过程的关键挑战。

1.2研究意义

**图神经网络在不同领域都得到了广泛的应用，高效的图神经网络建模加速技术为解决不同领域的关键问题提供了新的有效途径。**例如，电商网络中欺诈用户的检测[7]、自然语言处理中关系抽取[8]、社交网络中链路预测[9]和节点聚类[10]、生物医疗中药物副作用的预测[11]。我们选择以下几个典型应用进行介绍。（1）金融风险评估。金融行业内部存在大量风控需求，如反洗钱、防身份欺诈、防车险骗保、防金融欺诈、信用卡伪造交易套现等。图神经网络的拓展性和线上预测能力能够根据资金交易关系网络构建动态图模型，发现个体或群体的异常交易行为，对金融业务中的每一笔业务进行风险预测。（2）电商应用。伴随着电商用户的快速增长，平台上用户欺诈行为增多，包括刷单炒信、虚假评价、风险流量等，这些欺诈行为严重影响搜索排序算法的准确率，成为电商生态的重要威胁。阿里巴巴已经利用图神经网络进行欺诈用户的检测，取得了良好的效果。（3）自然语言处理。在自然语言处理中，包括多跳阅读[12]、实体识别[13]、关系抽取[8]以及文本分类[14]都可以运用图神经网络解决相关问题。（4）生物医疗。化合物是由原子和化学键构成的，它们天然就是一种图数据的形式，所以图神经网络在生物医疗领域应用特别广泛。包括新药物的发现[15]、化合物筛选[16]、蛋白质相互作用点检测[17]、以及疾病预测[18]。目前国内外已经有很多实验室研究图神经网络在医学方面的应用，这将是图神经网络最有价值的应用方向之一。

除了上述的应用领域外，在包括交通预测[1]、程序推断[19]、优化求解[20]等任务上，图神经网络都开始被人们使用。由于其可以在现实生活中常见的图数据上建模，并且能够通过卷积、注意力或消息传播等机制，将网络的拓扑结构和节点属性等信息用神经网络进行捕获和建模，图神经网络有广泛的应用前景与重要的研究意义。

1.3.国内外研究现状

近年来，图数据规模呈指数级增长，形成包含数十亿个多种类型的节点和数万亿条多种类型边的大规模图数据。单机模式下的图计算已经不适合目前数据的增长，传统的分布式数据处理平台比如MapReduce、Spark也出现网络和磁盘读写开销大、运算速度慢、处理效率极低的问题。由于图神经网络在图数据上有很强的特征挖掘能力，越来越多的图神经网络技术应用在大规模图数据上，并取得了相应的成果。

最近几年国内外各大互联网科技公司在大规模图数据处理与挖掘领域投入了很多资源进行布局，并从实际应用与业务场景出发落地了一系列基于大规模图数据的图神经网络建模技术与平台。AliGraph已部署在阿里巴巴电商平台上，以支持包括产品推荐和个性化搜索等各种业务场景。在具有4.9290亿个顶点，68.2亿条边和丰富属性的真实数据集上，AliGraph在构图速度上相较于PowerGraph提高了一个数量级。同时，AliGraph使用优化的缓存策略将训练速度提高了40%-50%[21]。腾讯开发了Plato图计算框架，应用在亿级的微信用户图数据上构建用户画像，Plato针对不同类型的图算法，提供了稀疏稠密自适应计算模式、共享内存计算模式和流水线计算模式等，还设计了良好的接口支持接入新的计算通信模式，并包含了图划分、图表示以及多层级计算通信协同调度等模块。亚马逊开源了DGL[22]大规模图神经网络计算框架，并基于此框架构建图神经网络用于推荐、欺诈检测等相关业务。DGL与多种深度学习框架深度集成，并实现了更加优秀的内存管理机制，支持在更大规模的图上训练模型，并基于图分区和图采样提供了分布式支持。Facebook通过自研的图表示学习分布式框架Pytorch BigGraph[23]，对大规模社交网络图数据进行图神经网络建模挖掘用户价值信息。Pytorch BigGraph支持在包含数十亿节点和数万亿边的大型网络上训练模型。该框架通过图分区、单机多线程、分批负采样等方式对分布式训练进行优化。

但目前国内外主流图神经网络建模技术中，没有基于大规模图数据特征，高效、自动地构建图神经网络模型的相关能力[24]，而基于图特征自动建模有助于提高建模准确度加速挖掘图数据潜在价值帮助相关应用的落地，所以基于大规模图数据的自动图神经网络建模加速技术的研究对于生产实践有着重要意义。

参考文献:

1. Cai H, Zheng V W, Chang K C C. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering*. 2018, 30(9): 1616-1637.
2. Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.
3. Chiang W L, Liu X, Si S, et al. Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks. *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019: 257-266.
4. Gao Y, Yang H, Zhang P, et al. GraphNAS: Graph neural architecture search with reinforcement learning. *Proceedings of International Joint Conference on Artificial Intelligence* (IJCAI). 2020: 1403-1409.
5. Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. *Proceedings of International Conference on Learning Representations*. 2017: 1–14.
6. Chiang W L, Liu X, Si S, et al. Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks. *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019: 257-266.
7. Wang H, Li Z, Huang J, et al. Collaboration based multi-label propagation for fraud detection. *Proceedings of International Joint Conference on Artificial Intelligence*. 2020: 2477-2483.
8. Nathani D, Chauhan J, Sharma C, et al. Learning attention-based embeddings for relation prediction in knowledge graphs. *Proceedings of Annual Meeting of the Association for Computational Linguistics*. 2019: 4710-4723.
9. Zhang M, Chen Y. Link prediction based on graph neural networks. *Proceedings of Advances in Neural Information Processing Systems*. 2018: 5165-5175.
10. Park H, Neville J. Exploiting interaction links for node classification with deep graph neural networks. *Proceedings of International Joint Conference on Artificial Intelligence*. 2019: 3223-3230.
11. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*. 2018, 34: i457–i466
12. Park H, Neville J. Exploiting interaction links for node classification with deep graph neural networks. *Proceedings of International Joint Conference on Artificial Intelligence*. 2019: 3223-3230.
13. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*. 2018, 34: i457–i466
14. Qiu L, Xiao Y, Qu Y, et al. Dynamically fused graph network for multi-hop reasoning. *Proceedings of Annual Meeting of the Association for Computational Linguistics*. 2019: 6140-6150.
15. Gui T, Zou Y, Zhang Q, et al. A lexicon-based graph neural network for Chinese ner. *Proceedings of the Conference on Empirical Methods in Natural Language Processing and International Joint Conference on Natural Language Processing* (EMNLP-IJCNLP). 2019: 1039-1049.
16. Hu L, Yang T, Shi C, et al. Heterogeneous graph attention networks for semi-supervised short text classification. *Proceedings of the Conference on Empirical Methods in Natural Language Processing and International Joint Conference on Natural Language Processing* (EMNLP-IJCNLP). 2019: 4823-4832.
17. Lin X, Quan Z, Wang Z J, et al. Kgnn: Knowledge graph neural network for drug-drug interaction prediction. *Proceedings of International Joint Conference on Artificial Intelligence*. 2020: 2739-2745.
18. De Cao N, Kipf T. MolGAN: An implicit generative model for small molecular graphs. *Proceedings of ICML workshop on Theoretical Foundations and Applications of Deep Generative Models*. 2018: 1-11.
19. Khalil, Elias, et al. Learning combinatorial optimization algorithms over graphs. *Proceedings of Advances in Neural Information Processing Systems*. 2017: 6348-6358.
20. Li Z, Chen Q, and Vladlen Koltun. Combinatorial optimization with graph convolutional networks and guided tree search. *Proceedings of Advances in Neural Information Processing Systems*. 2018: 539-548.
21. Shixun Huang, Yuchen Li, Zhifeng Bao, Zhao Li: Towards Efficient Motif-based Graph Partitioning: An Adaptive Sampling Approach. ICDE 2021
22. Zhao Li, Long Zhang, ChenyiLei, Xia Chen, Jianliang Gao, Jun Gao:. Attention with long-term interval-baseddeep sequential learning for recommendation[J]. Complexity, 2020, 2020.
23. Zhao Li, Haobo Wang, Donghui Ding, Shichang Hu, Zhen Zhang, WeiweiLiu, Jianliang Gao, Zhiqiang Zhang, Ji Zhang: Deep Interest-Shifting Networkwith Meta-Embeddings for Fresh Item Recommendation[J]. Complexity, 2020, 2020.
24. 高建良，张丽霞，《大规模图数据计算技术》，中南大学出版社，2018.

二、项目目标及研究内容

2.1.关键科学问题

**(1) 如何基于大规模图数据加速图神经网络训练**

实际生产环境中图数据规模基本在亿级，导致图神经网络在大规模图数据上的训练效率低下，甚至在有限的资源下无法进行训练。如何基于大规模图数据加速或实现单个图神经网络训练是拟解决的第一个关键科学问题。我们拟提出自适应计算资源的高效图采样技术，在保证采样子图结构特征与原始大规模图相似的情况，大幅度缩小用于训练的子图规模，从而实现图神经网络在大规模图上训练可行性并有效加速训练速度。

**(2) 如何基于大规模搜索空间加速图神经网络的建模过程**

图神经网络建模过程中最重要的一环是搜索图神经网络各个结构组件的组合，其搜索空间规模与图神经网络层数成指数关系。而在大规模图中训练一个已确定结构的图神经网络是非常耗时的过程，因此，在实际大规模图应用中无法遍历评估搜索空间中每一种图神经网络结构。如何在庞大的搜索空间中使用高效的搜索算法，自动找到最优图神经网络结构是拟解决的另一个关键科学问题。我们拟提出在搜索过程中向搜索算法引入图神经网络结构信息约束搜索方向，结合分布式计算能力与进化学习设计基于图特征自适应搜索算法，大幅提高图神经网络建模速度。

2.2研究目标

本项目的总目标为：从研究大规模图数据图神经网络模型出发，重点在大规模图数据采样技术与图神经网络结构并行搜索核法两个方面取得突破，并在不同的应用中运用这些技术，为大规模图神经网络提供一种有效的解决方案，从而推动相关应用领域的发展。本项目拟定研究目标为以下两点:

**(1)** **面向大规模图数据的图采样。**合理的缩小原始大规模图数据的规模，是在有限资源下实现图神经网络在大规模图上训练或者加速训练过程最直接最有效的方式。本项目将重点研究图采样过程中采样子图与原始大规模图结构特征相似性问题，保证采样子图与原始大规模图的结构特征具体相似性。同时，在权平衡率的情况下，设计合理的采样机制保证采样子图的多样性。最后，将研发内存评估算法评估采样子图的规模，使得采样算法能自适应的基于算力采样合适规模的子图为图神经网络训练提供可靠的数据支撑。

(2) 基于大规模搜索空间的图神经网络结构并行搜索与模型集成。超大空间的搜索问题，时间开销往往是巨大的，启发式搜索算法结合并行搜索方式是改进搜索效率的重要手段。本项目将基于进化学习利用图神经网络结构信息熵约束搜索方向，同时，在并行搜索的过程中，设计共享机制使得每个搜索器能充分共享搜索信息，大幅加速整个搜索过程。为了增强模型鲁棒性与准确性，本项目将设计模型集成机制，搜索出的图神经网络结构进行测试后将表现良好的模型进行集成，共同对新的样本进行预测，使用相应的集成策略进行结果判断。

2.3研究内容

在数据爆发式增长的背景下，如何自适应的针对大规模数据自动构建最优图神经网络结构已经开始得到工业界的关注，然而大规模图神经网络的建模加速面临着巨大挑战。本项目研究方案总体流程图如图2所示。本项目研究内容将从大规模图数据采样、图神经网络结构并行搜索两个角度切入。

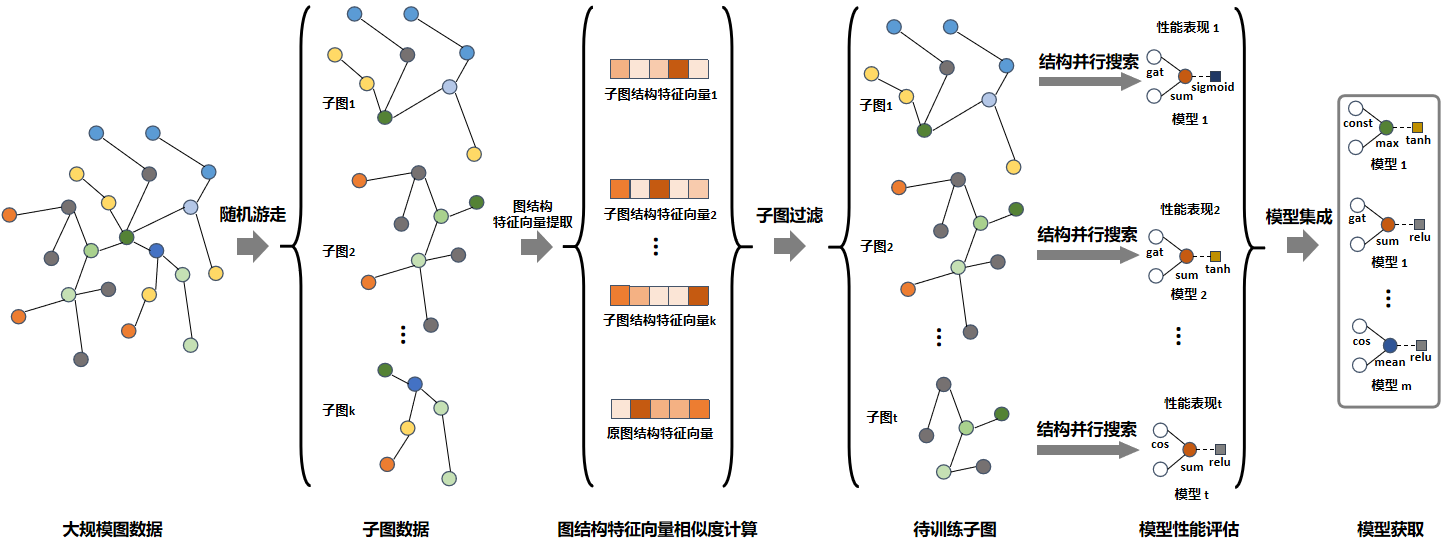


图 2 基于大规模图数据的图神经网络建模加速

**2.3.1.面向大规模图数据的图采样**

**随机游走采样。**使用内存测试程序自动评估单台计算设备内存空间,基于评估内存空间计算子图规模s,根据大规模图数据选择不同的节点为起始节点，使用随机游走算法采样k个规模为s的子图，对子图节点重复率进行评估，保证图采样的全面性。

**基于图结构特征向量的子图过滤。**为了保证采样子图与原始大规模图具有结构相似性，使用高效图结构特征向量提取算法NetLSD，提取原始大规模图与k个子图的结构特征向量，基于马特恩协方差函数评估k个子图与原始大规模图结构特征向量的相似性，选取相似性最高的t个子图代表原始大规模图为图神经网络的结构搜索与模型训练提供数据支撑。

2.3.2.基于大规模搜索空间的图神经网络结构并行搜索与模型集成

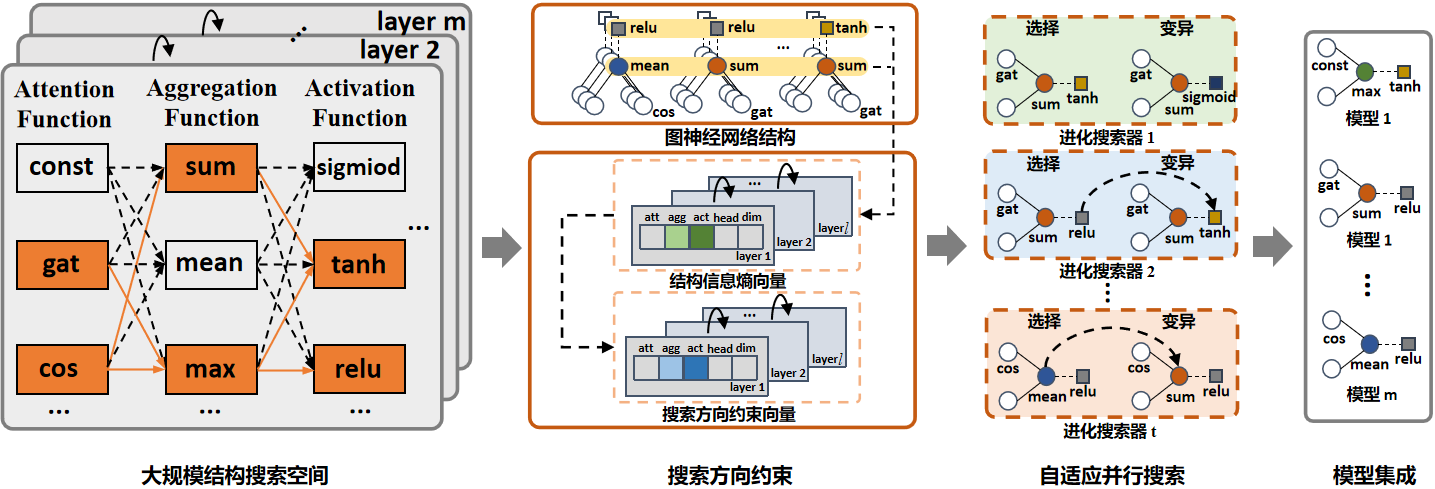


图 3 图神经网络结构结构并行搜索

**搜索方向约束。**经过对图神经网络结构深入分析与研究，基于给定的图数据特征分布，性能表现好的图神经网络结构具有相似性，我们将使用图神经网络的结构信息熵约束搜索方向，缩小超大规模搜索空间范围。

**自适应并行搜索。**基于进化学习设计基于图数据特征的并行自适应搜索算法，使用多个搜索器在t个子图中并行地进行结构搜索，并设计高效的信息共享机制，使得每个搜索器能使用结构信息熵约束搜索方向，同时，充分利用其它搜索器的搜索结果大幅加速搜索过程。

**模型集成**。基于图神经网络结构性能表现，选择top m个图神经网络模型进行集成，使用投票机制实现对输入样本的最终预测，增加模型的鲁棒性与准确性。图神经网络结构并行搜索流程图如图3所示

三、申报单位研究基础

3.1.项目申请团队基本情况

本项目团队是由中南大学计算机学院高建良教授牵头，张祖平教授、贺建飚副教授参与，以及计算机学院博士研究生和硕士研究生组成。本项目组近年来一直着力于研究图数据处理相关理论和应用，在大规模图数据计算、高效图处理算法、图神经网络应用等方面取得了一系列成果。  
 为了更加高效的处理大规模图数据，我们提出了分布式图查询算法、增量的图匹配算法、以及Top-K的子图查询算法。在最近的研究工作中，我们将重点对图神经网络进行了深入研究和应用。例如，我们提出了一种TypedGAT图神经网络实现了跨语言实体对齐。图数据隐含了复杂多样的特征，多个图数据关联的背后往往蕴含着单个图数据所不具备的知识，我们通过在多种图数据进行特征挖掘，研究跨语言知识图谱的实体对齐。在生物网络中，我们提出了基于图神经网络的数据挖掘方法，例如多模态数据融合预测癌症患者预期寿命、蛋白质相互作用网络对齐方法。在社交网络应用中，我们提出了K-匿名的图结构保护方法。在金融股票应用中，我们提出了动态图神经网络进行股票预测。

**团队代表性研究成果**：

1. Jianliang Gao, Ling Tian, Tengfei Lv, Jianxin Wang, Bo Song, Xiaohua Hu: Protein2Vec: Aligning Multiple PPI Networks with Representation Learning. IEEE-ACM Transactions on Computational Biology and Bioinformatics. 18(1): 240-249 (2021).
2. Zhao Li, Yuying Xing, Jiaming Huang, Haobo Wang, Jianliang Gao, Guoxian Yu: Large-scale online multi-view graph neural network and applications. Future Gener. Comput. Syst. 116: 145-155 (2021)
3. Gao J, Liu P, Kang X, et al. PRS: Parallel Relaxation Simulation for Massive Graphs. Computer Journal, 2016, 59(6), pp. 848-860.
4. Zhang L, and Gao J. Incremental Graph Pattern Matching Algorithm for Big Graph Data. Scientific Programming, 2018, 2018(6749561), pp. 1-8.
5. 张丽霞, 王伟平, 高建良,王建新. 面向模式图变化的增量图模式匹配. 软件学报, 2015. 26(11), pp. 2964-2980.
6. Gao J, Song B, Liu P, et al. Parallel top-k subgraph query in massive graphs: Computing from the perspective of single vertex. IEEE International Conference on Big Data, 2016, pp. 636-645,
7. Xiong F, and Gao J. Entity Alignment for Cross-lingual Knowledge Graph with Graph Convolutional Networks. The 28th International Joint Conference on Artificial Intelligence, 2019, pp. 6480-6481.
8. Gao J, Lyu T, Xiong F, et al. MGNN: A Multimodal Graph Neural Network for Predicting the Survival of Cancer Patients. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020: 1697-1700.
9. Gao J, Song B, Ke W, et al, BalanceAli: Multiple PPI Network Alignment With Balanced High Coverage and Consistency, IEEE Transactions on NanoBioscience, 2017, 16(5), pp. 333-340.
10. Song B, Gao J, Du H, et al. Aligning Multiple PPI Networks with Representation Learning on Networks. IEEE International Conference on Bioinformatics and Biomedicine, 2018, pp.136-141.
11. Gao J, Song B, Z. Chen, et al. Counter deanonymization query: H-index based K-anonymization privacy protection for social networks. The 40th International Conference on Research and Development in Information Retrieval, 2017, pp. 809-812 .
12. Gao J, Ping Q, and Wang J. Resisting re-identification mining on social graph data. World Wide Web, 2018, 21(6), 1759–1771.
13. Gao J, Wang J, He J, et al. Against Signed Graph Deanonymization Attacks on Social Networks. International Journal of Parallel Programming, 2017,47(4),pp.725-739.
14. Ying X, Xu C, Gao J, et al. Time-aware Graph Relational Attention Network for Stock Recommendation. ACM International Conference on Information & Knowledge Management (CIKM),2020:2281–2284.
15. Abdelaziz M, Zhang Z. Few-shot learning with saliency maps as additional visual information. Multimedia Tools and Applications, 2021, 80(7): 10491-10508.
16. Chen H, Zhang Z, Zhang J. In silico drug repositioning based on the integration of chemical, genomic and pharmacological spaces. BMC bioinformatics, 2021, 22(1): 1-12.
17. Peng L, Chen Z, Chen T, et al. Prediction of the Age at Onset of Spinocerebellar Ataxia Type 3 with Machine Learning. Movement Disorders, 2021, 36(1): 216-224.
18. Long Z, Zhang Z. Optimization and Deployment of Vehicle Trajectory Prediction Scheme Based on Real-Time ANPR Traffic Big Data. International Conference on Artificial Intelligence and Security. Springer, Cham, 2020: 74-85.
19. Chen H, Guo R, Li G, et al. Comparative analysis of similarity measurements in miRNAs with applications to miRNA-disease association predictions. BMC bioinformatics, 2020, 21: 1-14.

3.2.项目申请牵头人基本情况

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，目前担任阿里巴巴集团项目合作专家。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] **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)

[4] 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)

[5] **Jianliang Gao**, Ling Tian, Tengfei Lv, Jianxin Wang, Bo Song, Xiaohua Hu: Protein2Vec: Aligning Multiple PPI Networks with Representation Learning. IEEE-ACM Transactions on Computational Biology and Bioinformatics. 18(1): 240-249 (2021).. (SCI)

[6] 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)

[7] **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)

[8] Jun Gao, **Jianliang Gao\***, A Similarity Measurement Method Based on Graph Kernel for Disconnected Graphs. IJCAI 2019. (CCF A类会议)

[9] Fan Xiong, **Jianliang Gao\***, Entity Alignment for Cross-lingual Knowledge Graph with Graph Convolutional Networks. IJCAI 2019. (CCF A类会议)

[10] Hongliang Du, Zhiyi Jiang, **Jianliang Gao\***, Who is Who: Name Disambiguation in Large-Scale Scientific Literature. ICDM Workshops 2019: 1037-1044.

[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.

四、进度安排

为了保证项目按照计划顺利实施，基于研究目标将研究内容进行拆解细化，如表1所示，整个研发过程分为六个部分，1.模型比较2.软件架构设计3.核心算法研发4.前端开发5.后端开发6.准备结题。

|  |  |  |
| --- | --- | --- |
| **研发** | **测试** | **时间** |
| 模型比较 | 1.模型比较评估 2.技术组件效率与生态评估 | 2021年04月1日 2021年06月31日 |
| 软件架构设计 1.前后端交互功能设计 2.前端页面设计 3.数据库设计 4.分布式服务架构设计 | 1.功能完备性验证 2.架构高内聚低耦合性验证 | 2021年07月1日 2021年12月31日 |
| 核心算法研发—大规模图采样算法 1.内存容量评估算法研发 2.基于大规模图采样的随机游走算法与采样策略研发 3.NetLSD图结构特征向量提取算法研发 | 1.算法科学性验证 2.算法效率验证 3.算法兼容性验证 4.算法扩展性验证 | 2022年1月1日 2022年5月31日 |
| 核心算法研发—基于进化学习的并行图神经网络结构搜索 1.基于结构信息熵变异的进化搜索算法研发 2.搜索算法并行机制消,息共享机制设计 3.图神经网络搜索空间设计 4.搜索算法反馈机制模块研发 5.不同类型图任务的训练、验证、测试模块研发 | 1.算法科学性验证 2.算法效率验证 3.算法兼容性验证  4.算法扩展性验证 | 2022年6月1日 2022年10月31日 |
| 前端开发 1.前端页面开发 2.前后端交互逻辑开发 | 功能测试 | 2022年11月1日 2022年12月31日 |
| 后端开发 1.分布式服务架构开发 2.数据库开发 | 1.分布式服务功能测试 2.数据库功能测试 | 2023年1月1日 2021年2月30日 |
| 总结成果，准备结题 | 结题报告材料 | 2021年3月1日 2022年3月30日 |

表1.研发计划表

五、项目实施机制、保障措施及风险分析

5.1 项目实施机制

本项目建立项目负责人、项目专家组和课题负责人联席会议制度，并实行过程动态管理模式及项目负责人、课题负责人负责制，按照“项目—课题”分级管理。全体研究人员分工协作，责任落实到人，协同攻关研究。项目负责人全面负责本项目的研究目标、内容、任务的实施工作，组织本项目各课题间开展学术交流，检查项目研究工作进展，负责项目中期检查评估汇报和验收报告，充分调动项目参加人员的积极性和创造性，并及时向项目主管部门汇报课题实施和进展情况。

成立项目专家组，由外聘专家组成，专家组对项目和课题研究提供技术指导和咨询，为项目和课题负责人做出决策提供建议，参与项目和课题实施过程中的监督检查、审核、评估与验收等工作。组织学术研讨与交流，对项目和课题的关键技术问题、阶段研究成果及最终研究成果等进行技术咨询和把关。

本项目实行季度汇报评审制度，通过课题汇报和专家组评定，采取激励管理政策，鼓励创新，鼓励项目参与者深入项目，鼓励撰写高质量、高水平的研究论文，鼓励积极开展研究获取技术专利，对研究工作进展较好或取得重大进展的专题，增加研究经费支持力度。

在项目组织实施小组和项目专家组的领导下，成立课题负责人与各子课题负责人组成的课题实施小组，对课题进行组织和管理。课题实施小组，主要负责课题总体的宏观指导和对重大问题做出决策，采取课题承担单位管理的方式，课题承担单位负责组织课题的条件保障、具体实施和日常管理，及时发现和报告课题实施过程中出现的问题，按要求编报课题年度计划、计划执行情况和统计报表、科技攻关成果信息及经费预算等信息资料，并及时报项目组织单位。

5.2 保障措施

本项目依托湖南省科技厅和中南大学负责项目日常管理和提供条件保障，从行政组织和支撑条件各方面创造良好的研究环境，确保项目的顺利实施；审核汇总项目的经费预算和年度经费预算，检查、监督项目经费的管理和使用情况；了解项目的执行情况，审查项目专家组提出的年度报告和结题总结报告；负责项目的成果管理。

本项目在机制保障方面，成立组织实施小组和项目专家组的双轨制的组织模式。组织实施小组保障对专项总体的宏观指导和对重大问题的决策。项目专家组负责本项目的“顶层设计”和实施过程中的技术咨询指导，把握项目和课题的总体目标及研发方向。项目负责人全面负责项目的研究方向、内容和进展，协调解决项目中遇到的重大问题；各课题组长负责和协调各课题的研究工作，保证课题的按计划进展。在开展和推进研究工作中，项目组既会充分发挥各课题组的特色，也将不断围绕总体目标协同攻关，保证项目的顺利实施和目标的圆满实现。

项目负责人及团队在图计算、深度学习、和图神经网络等相关领域积累了丰富的研究基础和工作基础，拥有实力雄厚的研发队伍、良好的科研环境和实验条件，课题组成员具有扎实的基础知识、丰富的相关研究和实践经验以及良好的科研素质，为本项目的研究和实施提供了资源保障。

5.3 风险分析

1.技术风险。研发开发过程中可能出现难度较大的技术攻关难题。围绕大规模图神经网络模型研究，开展学术合作和交流。定期召开课题研讨会、建立权威专家技术攻坚小组，促建多学科和多领域的交叉融合，积极开展合作，引进吸收经验。

2.财务风险。项目负责人严格执行财务制度，控制资金去向和使用情况，确保财务透明，无财务风险。

六、其它需要说明的事项

无。