**北 京 邮 电 大 学**

**本科毕业设计（论文）中期进展情况检查表**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 学院 | | | 计算机学院 | 专业 | | 计算机科学与技术 | | 班级 | 304 |
| 学生姓名 | | | 陈童 | 学号 | | 2019211283 | | 班内序号 | 18 |
| 指导教师姓名 | | | 杨成 | 所在单位 | | 北京邮电大学计算机学院 | | 职称 | 副教授 |
| 设计（论文）题目 | | | 基于模型增强的异质图对比学习算法设计与实现 | | | | | | |
| 目前已完成任务 | 主要内容: (毕业设计（论文）进展情况，字数一般不少于1000字)  毕业设计主要步骤如下：首先在大四秋季学期学期结束前完成前期文献调研，选择算法和方法。在大四春季学期结束前完成实验设计、实验算法设计，并完成整体实验的设计构建任务。在2023年3月前完成实验环境的编写，开展实验验证，收集分析实验数据，并针对结果对模型进行调整和改善。在2023年5月前完成系统的综合测试，完成毕业设计论文的撰写，完成答辩。  目前已经完成了文献调研、算法的选择和设计、机器学习模型搭建和初步实验验证，收集实验数据和初步的数据分析工作。  前期调研方面，在调查现有的图对比学习以及异构图机器学习相关研究后，决定使用图卷积神经网络（Graph Convolutional Network，GCN）来提取和处理输入图数据的特征。GCN是一种用于处理图像数据的深度学习模型，它可以通过建模图像的拓扑结构来提取图像的空间特征，因此非常适合处理具有复杂结构的图像数据。此外，在前期调研中，我发现图对比损失（Graph Contrastive Loss）来比较不同图数据之间的相似性，有利于增加提取特征的多样性和过滤无关的特征。图对比损失是一种用于图像对比学习的损失函数，它可以将同类图像的特征向量拉近，不同类图像的特征向量拉远，从而增强图像分类的准确性。最后，我使用模型增强技术（Model Augmentation）来优化算法性能。模型增强是一种用于提高模型性能的技术，它可以通过增加训练数据、调整模型超参数、改进模型结构等方法来提高模型的泛化能力和鲁棒性。在本次研究中，我选择以GCN作为基础模型，并对不同的输入数据使用不同结构的GCN处理提取特征。值得注意的是，所有的GCN网络将共享参数。此外，在我进行前期文献调研时，我发现元路径（metapath）在图对比学习领域中有着重要的应用。元路径是一个由节点类型序列构成的路径，可用于在异构信息网络（Heterogeneous Information Network，HIN）中捕获节点之间的关联信息。在HIN中，不同类型的节点具有不同的语义含义，而元路径可以帮助我们刻画节点之间的语义关联，并设计有效的图对比学习算法。具体来说，元路径可以用于定义图中的节点相似性度量，从而引导图对比学习算法在节点嵌入空间中进行有效的优化。通过对元路径的定义和选择，我们可以更好地利用节点之间的关联信息，并提高图对比学习算法的准确性和泛化能力。  在对文献调研后，我开始进行算法设计。我选择使用GCN作为基础的编码器（Encoder）。在使用元路径方法对图数据进行多样化提取后，会生成若干元路径数据作为模型的输入。针对每一个输入，都将有一个对应的编码器对其进行处理。损失函数InfoLoss将会计算所有输出的交叉熵。在模型预测阶段，则是将所有编码器的结果取平均值后，作为整个模型的输出。  在模型测试时，我使用了四个数据集，分别是Freebase，Acm，Aminer和Dblp。在初步的训练中，我的模型实现了不输于基准模型的性能。 | | | | | | | | |
|  | | | | | | | | |
| 尚需完成的任务 | 接下来，我将进一步对初步测试的实验结果进行分析，并在此基础上对模型进行调整，以达到更好的性能。并且我将对现有成果进行总结，对数据进行更细致的分析和处理，准备毕业论文的撰写和答辩工作。 | | | | | | | | |
| 能否按期完成设计（论文）是 | | | | | | | | |
| 存在问题和解决办法 | 存  在  问  题 | 在进行毕业设计的实验过程中，我遇到了两个问题。第一个问题是在搭建Pytorch实验环境时出现了失败的情况，导致无法正常跑通实验。第二个问题则是在寻找网上的异质图数据集时，发现数量较少且缺少详细的文档说明。 | | | | | | | |
| 拟  采  取  的  办  法 | 对于第一个问题，经过在网上搜寻资料发现，问题出现在Pytorch\_geometric 安装包上。Conda和Pip方法的自动安装并不完善，容易出现问题。最终我选择在官网上通过下载指定的，符合设备和Cuda版本号的whl包手动安装解决。  第二个问题，，由于缺少文档说明，我无法快速了解数据集的具体情况和使用方法。为了解决这个问题，我首先尝试了在数据集官网上查找相关文档，但结果并不理想。最后，我发现有一篇论文在网上较好的公开了自己的代码和数据集，我选择参考他的数据集使用方法，实现了数据集的正常使用 | | | | | | | |
| 指导教师签字 | |  | | | 日期 | | 年 月 日 | | |
| 检查小组意见 | | 负责人签字： 年 月 日 | | | | | | | |

注：可根据长度加页。