**北京大学本科毕业论文中期审核表**

|  |  |  |  |
| --- | --- | --- | --- |
| 学生姓名 | 王宇哲 | 学号 | 1800011828 |
| 院系 | 化学与分子工程学院 | 专业 | 化学 |
| 指导教师 | 张长胜 | 职称 | 副研究员 |
| 论文中文题目 | 基于图神经网络的蛋白质主链结构优化 | | |
| 论文英文题目 | Optimization of Protein Backbone Structure Through Graph Neural Network | | |
| 毕业论文进展（含阶段性成果；注明累计研究工作时间） | 毕业论文的主要进展和阶段性成果如下：   1. 收集并处理数据集。使用PISCES服务器从PDB中筛选出14606个结构，编写脚本去除无效结构（末端残基缺失等）后得到14082个结构，作为模型训练所使用的数据集。 2. 编写程序实现图神经网络模型，并对模型进行训练。编程实现模型所需的必要组件及图神经网络模型，在前述数据集上成功完成训练，模型代码及工作进度见<https://github.com/Division-Bell-CCME18/ProteinBackboneDesign>。 3. 模型改进与调参。对上述模型中图的构建进行调整，显著提升了模型的表现；通过反复实验调整图神经网络模型的各参数（降噪神经网络的噪声强度σ、学习率、学习率衰减、batch size等），对模型的表现进行优化。   累计研究工作时间约为330 h（大四上学期及寒假约50 h；本学期每周约40 h，共7周）。 | | |
| 目前存在的主要问题 | 目前存在的主要问题如下：   1. 图的构建可能尚存在节点选取不够合理、边的物理意义不够明确等问题，需要对节点之间的连接方式加以改进，考虑如何把蛋白质主链氢键相互作用更恰当地加入图的构建中。 2. 需要继续优化模型表现，通过后续调参，实现模型性能的提升，目前模型的表现（train loss, validation loss等）仍不能令人满意。 3. 尚未对模型在具体蛋白上的表现进行测试，需要尽快完成模型用于具体蛋白主链构象优化的脚本，并进行适当的可视化处理，对模型的表现加以具体评估。 | | |
| 下阶段工作计划 | 下阶段的工作计划如下：   1. 进一步修改并优化模型，考虑对图的构建进行深度修改，并查阅相关文献，通过有目的性地调参提升模型性能。 2. 尽快完成模型对具体蛋白主链构象进行优化的脚本及可视化处理，测试模型在具体蛋白上的表现。 3. 着手撰写毕业论文，考虑论文的整体结构，拟定论文的大纲及写作计划。 | | |
| 指导教师评价 | 指导教师签字：  2022年\_4\_月\_11\_日 | | |
| 院系意见 | （公章）  2022年\_4\_月\_12\_日 | | |