

通过上一步的运行，我们得到了原始的测试数据。将原始测试数据经初步处理和整合后得到的结果如下：

表 5.4 不同模型在不同数据集上的准确率

	Citeseer	Cora	Pubmed
MLP	0.475, 0.0212	0.497, 0.0213	0.734, 0.0248
Spectral CNN	0.476, 0.0237	0.497, 0.0243	0.734, 0.0201
FirstCheb	0.600, 0.0255	0.581, 0.0352	0.713, 0.0493
DCNN	0.640, 0.0213	0.661, 0.1374	- -
GCN	0.648, 0.0280	0.778, 0.0238	0.746, 0.0735
GAT	0.650, 0.0207	0.790, 0.0267	0.764, 0.0509
Graphsage-Mean	0.622, 0.0262	0.609, 0.0427	0.604, 0.0408
Graphsage-MeanPool	0.592, 0.0330	0.719, 0.0454	0.770, 0.0484
Graphsage-Maxpool	0.577, 0.0208	0.706, 0.0179	0.768, 0.0342

上图为本文选择的图神经网络模型在不同数据集上测试的结果，其中每格数据的第一项为测试得到的准确率，第二项为模型在不同的数据集分割上测试的得到结果的标准差。需要注意的是，本文在测试过程中使用了 10 个不同的数据集分割，在每个数据集分割上进行了 3 次测试。

标准差的计算方法：首先对每个数据集上进行的三次测试取平均值，对 10 个不同的数据集分割测试得到的平均值计算标准差。因此，每个数据集划分上测试的次数可能会对标准差的值产生影响。故本文给出的标准差仅起到对比的作用。

为了更加直观的对比不同模型的准确率差异，将上表中的数据绘制成图像。其中，图像中竖线的中点表示该神经网络在本数据集上的平均准确率，上限和下限分别为图神经网络模型的预测准确率与标准差的和和差。通过下列图像，我们可以直观的比较出不同图神经网络模型的性能和稳定性之间的差异。

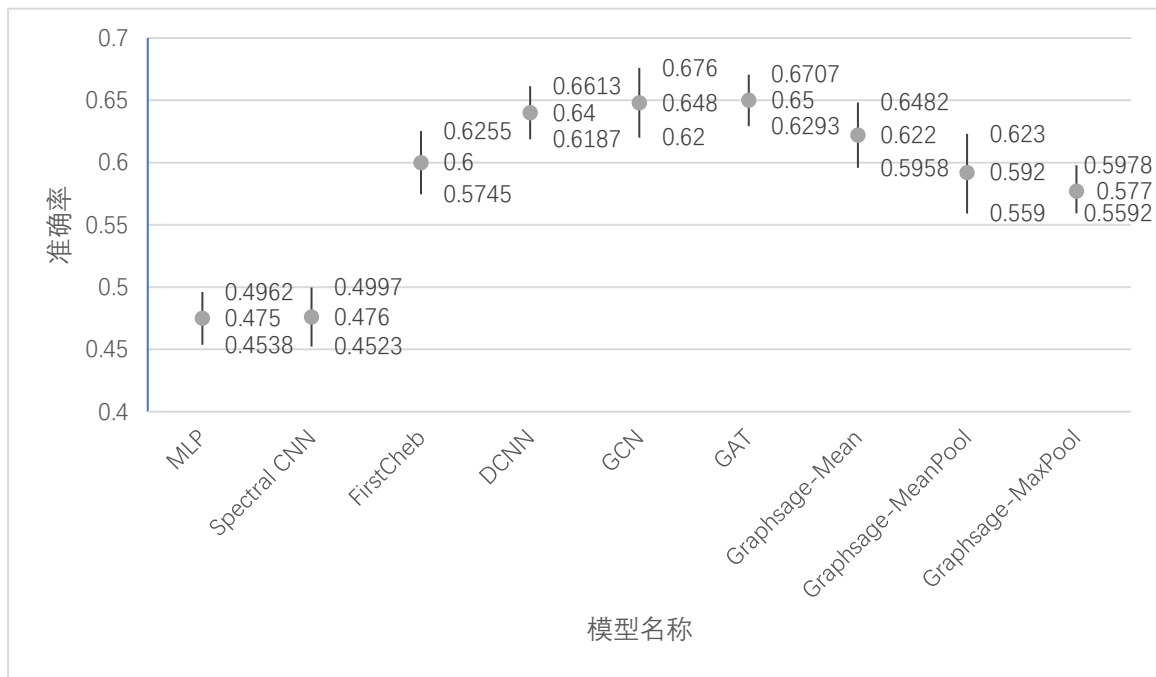


图 5.6 不同模型在 Citeseer 数据集上的准确率差异

图中可以看到，图神经网络模型表现出了远超过基线模型的准确率，并且拥有较高的稳定性。其中 GCN 和 GAT 两个模型准确率十分接近且高出其他模型。GraphSage 模型相对于其他图神经网络模型来说，三个聚集函数模型都没有得到最好的效果。

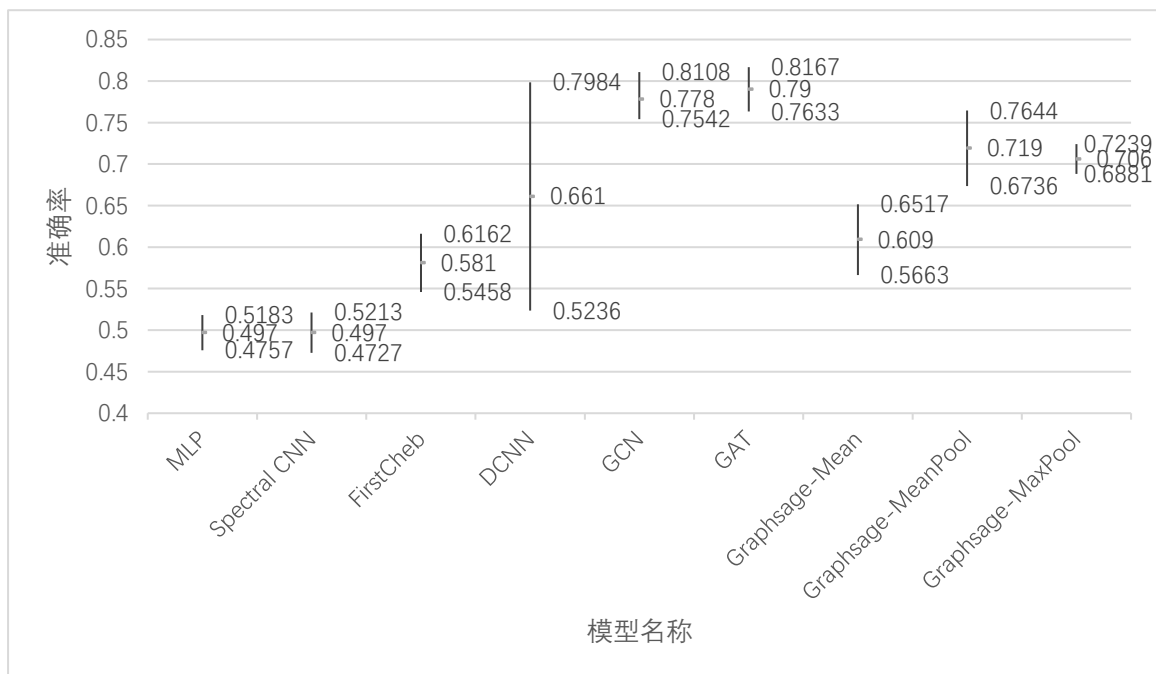


图 5.7 不同模型在 Cora 数据集上的准确率差异

从图中可以看出，在 Cora 数据集上，图神经网络模型仍然取得了远超过基

线模型的准确率。其中 GCN 和 GAT 模型仍然取得了最好的效果。在图神经网络模型中, FirstCheb 和 GraphSage-Mean 函数中准确率较低。需要注意的是, DCNN 模型在 Cora 数据集上稳定性较差。

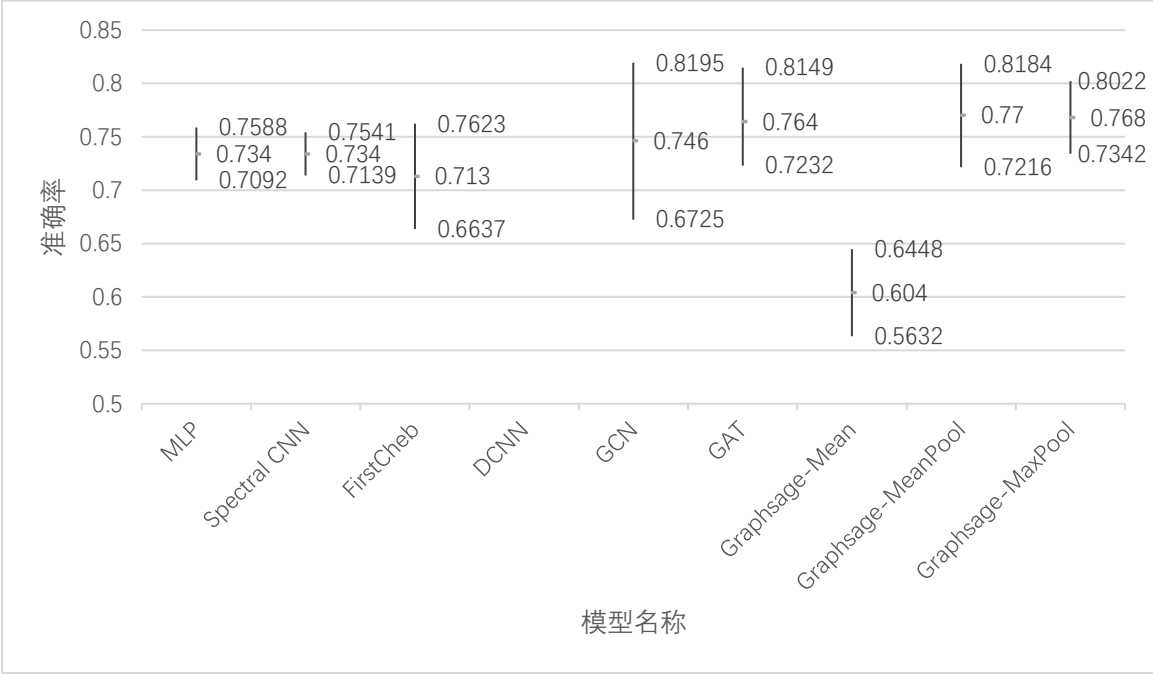


图 5.8 不同神经网络模型在 Pubmed 数据集上的差异

需要注意的是, 本文没有在 Pubmed 数据集上测试 DCNN, 因 DCNN 需要的显存高于其他模型, 故在 Pubmed 数据集上运行 DCNN 时, 出现了显存溢出的状态。考虑到 DCNN 模型需要的变量和输入信息, 想对比于其他模型来说, DCNN 模型需要的显存较高。在后续分析过程中, 我们将会考虑到模型内存占用量和模型编写复杂度的问题。

通过对最终得到的数据进行分析, 我们发现。除 Spectral CNN 外所有的图神经网络模型相对于我们选定的基线模型, 都得到了很好的效果。而且在节点种类和每个节点特征数量越多的图上, 图神经网络的效果就会越好。DCNN 在测试的模型中内存占用量最好, DCNN 在 Pubmed 数据集上进行测试时, 出现了显存溢出的问题。综合以上测试得到的结果 GCN 和 GAT 模型的效果较好。

除了模型的准确率, 我们还记录了在模型测试时使用的时间。以下时间为在一次运行中使用的时间, 即从向模型中输入数据到模型给出测试的结果需要的时间。

不同模型在不同数据集上测试需要的时间如下, 其中第一项为模型在单次测试过程中使用的时间, 第二项为多次测试过程中时间的标准差:

表 5.5 不同模型在不同数据集上需要的测试时间

	Citeseer	Cora	Pubmed
MLP	0.02170, 0.00452	0.03369, 0.00443	0.10451, 0.00505
Spectral CNN	0.05290, 0.00395	0.04869, 0.00379	1.19301, 0.00764
FirstCheb	0.02686, 0.00667	0.04615, 0.00659	0.13131, 0.00923
DCNN	0.10842, 0.00632	0.08410, 0.00707	-, -
GCN	0.02353, 0.00510	0.03697, 0.00497	0.10884, 0.00653
GAT	0.08373, 0.02750	0.19361, 0.03673	0.43763, 0.04515
Graphsage-Mean	0.02779, 0.00630	0.04853, 0.00743	0.13427, 0.00950
Graphsage-MeanPool	0.03069, 0.00755	0.05731, 0.00925	0.15402, 0.01158
Graphsage-Maxpool	0.02815, 0.00805	0.05649, 0.00927	0.14985, 0.01528

为了更加直观的比较出不同模型运行需要的时间，我们将模型测试需要的时间用柱状图表示。由于同一模型在 Pubmed 和 Citeseer 与 Cora 区别较大，在同一图上表示会导致无法直观观察前两个数据集的对比结果，故使用两个柱状图来进行表示。

从表中可以看出，模型测试时间的标准差差距较小。故在下图中没有表示出模型测试的标准差。

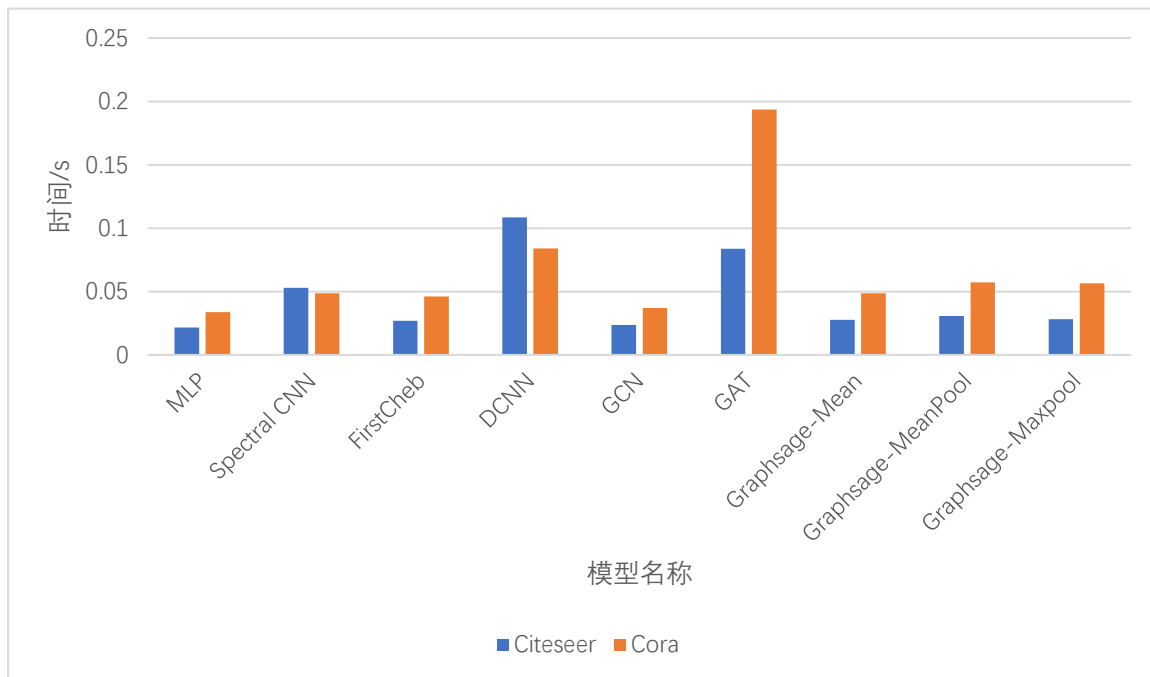


图 5.9 不同模型在 Citeseer 和 Cora 测试 需要的时间

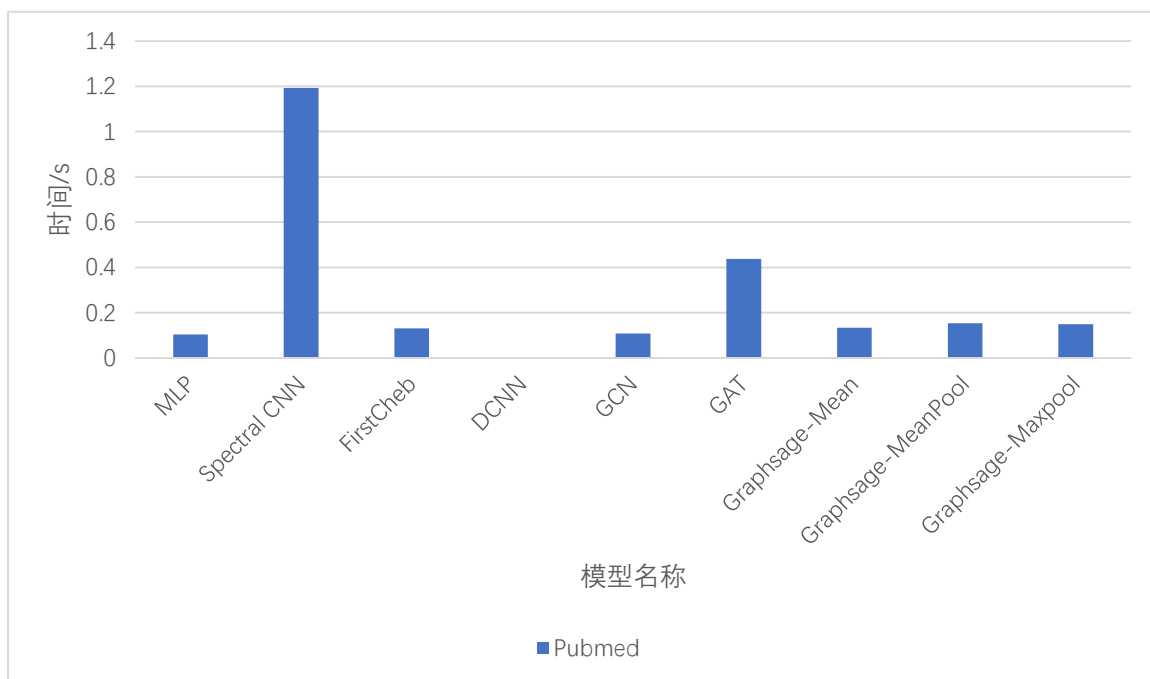


图 5.10 不同模型在 Pubmed 测试需要的时间

从上述两幅图中可以看出，在图中节点数量较小时，DCNN 和 GAT 使用的时间较长，其余模型的运行需要的时间较短。表现最好的是 GCN 模型，测试需要的时间接近基线模型，其余模型运行需要的时相近，稍高于 GCN。

当训练使用的图的节点数量较高时，DCNN 由于内存原因无法进行测试，表现最差的模型时 Spectral CNN，使用时间超过 1 秒。其次是 GAT，使用的时间在

4 秒到 5 秒之间。最快的模型依旧是 GCN，使用的时间仅稍稍高于基线模型 MLP。其余模型使用的时间均稍高于 GCN，且用时相近。

需要注意的是，在模型测试中，我们没有计算数据预处理需要的时间。大部分模型仅仅对原始数据的输入形式进行一些简单的修改，用时极短且可以忽略不计。DCNN 需要计算数据邻接矩阵的幂级数，用时稍长。预处理使用时间最多的是 Spcetral CNN，该模型在预处理时需要对邻接矩阵进行特征值分解，时间复杂度较高。在本实验中，数据的预处理使用 numpy 中的的函数并且在 CPU 中进行，因此计算需要的时间远远高于相同任务在 CPU 中运算需要的时间。本系统没有对预处理需要的时间进行严格的计时，但在 CPU 中对 Pubmed 的邻接矩阵进行特征值分解需要数分钟的时间，远远高于其他模型数据预处理需要的时间和运行需要的时间。因此，Spectral CNN 模型运行需要的时间远高于图中给出的部分。而且 Spectral CNN 准确率较低，故不建议使用 Spectral CNN。

### 模型准确率与训练集大小

本部分的内容是测试训练集大小对于图神经网络模型准确率的影响。在测试过程中，使用在上一步得到的超参数，对模型在 Citeseer 数据集上进行训练。模型的数据集被划分成三部分，训练集，验证集和测试集。在训练过程中，使用大小固定为 500 的验证集，其余部分被划分为训练集和测试集。本次测试在 0 个节点和 1400 个节点之间选择若干离散值进行测试。其中，训练集，验证集和测试集仍然按照节点顺序进行划分。这种划分方式得到的验证集部分不会完全相同，进一步增加了数据的变化，使得测试结果更加公平。

在本测试中得到的结果如下：

表 5.6 模型准确率与训练集大小之间的关系

	MLP	Sepectral CNN	FirstCheb	GCN	GAT
0	0.191	0.191	0.191	0.191	0.191
5	0.240	0.234	0.279	0.262	0.209
10	0.275	0.294	0.327	0.291	0.275
15	0.290	0.270	0.346	0.341	0.346
20	0.338	0.332	0.434	0.426	0.379
30	0.370	0.368	0.471	0.462	0.468

40	0.367	0.375	0.502	0.495	0.491
60	0.403	0.391	0.518	0.554	0.545
100	0.446	0.432	0.575	0.592	0.608
140	0.456	0.454	0.597	0.627	0.632
180	0.460	0.462	0.607	0.621	0.620
220	0.483	0.484	0.604	0.649	0.643
260	0.489	0.497	0.611	0.653	0.653
300	0.498	0.488	0.632	0.673	0.665
400	0.513	0.499	0.617	0.681	0.690
500	0.533	0.525	0.611	0.698	0.700
600	0.537	0.538	0.612	0.700	0.704

续表 5.6-1

800	0.540	0.550	0.603	0.721	0.719
1000	0.574	0.568	0.615	0.709	0.725
1200	0.571	0.576	0.620	0.732	0.732
1400	0.592	0.587	0.627	0.731	0.740

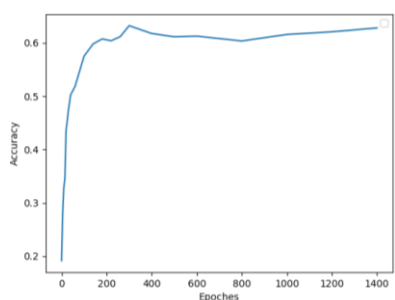
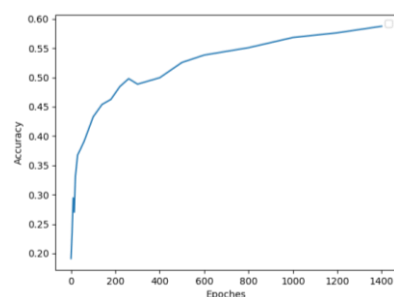
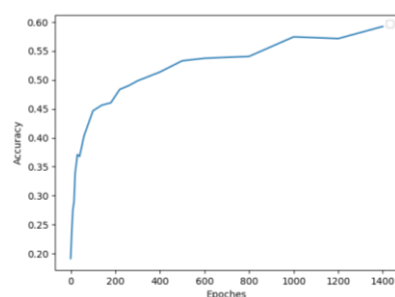
续表 5.6-2

	DCNN	Graphsage-mean	Graphsage-meanpool	Graphsage-maxpool
0	0.191	0.191	0.191	0.191
5	0.264	0.282	0.241	0.216
10	0.289	0.336	0.296	0.258
15	0.300	0.381	0.276	0.298
20	0.381	0.437	0.344	0.345
30	0.463	0.478	0.358	0.380
40	0.477	0.498	0.412	0.426
60	0.504	0.536	0.457	0.455
100	0.562	0.586	0.536	0.518
140	0.546	0.601	0.564	0.525
180	0.623	0.616	0.588	0.547

220	0.616	0.613	0.588	0.564
260	0.643	0.617	0.603	0.580
300	0.651	0.624	0.632	0.599
400	0.669	0.633	0.639	0.615
500	0.641	0.639	0.644	0.614
600	0.694	0.627	0.658	0.632
800	0.675	0.621	0.686	0.656
1000	0.680	0.621	0.698	0.647
1200	0.710	0.627	0.712	0.672
1400	0.704	0.633	0.709	0.684

为了更好的观察测试结果，直观的观察模型预测准确率随训练集数量变化的过程，我们绘制了训练集大小-模型准确率折线图。

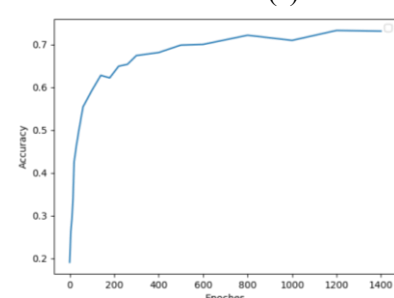
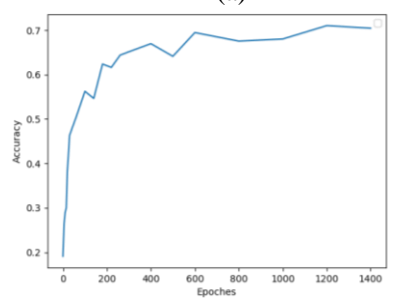
训练集大小-模型准确率折线图如下：



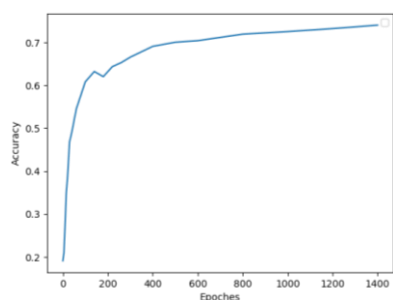
(a)

(b)

(c)



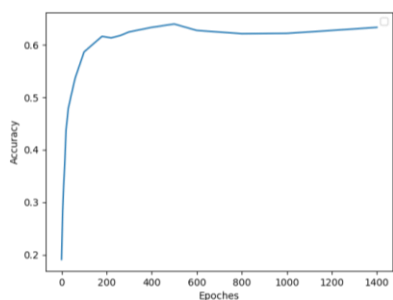




(d)

(e)

(f)



(g)

(h)

(i)

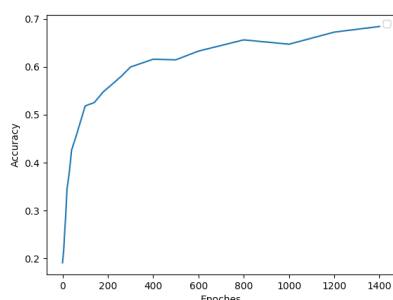
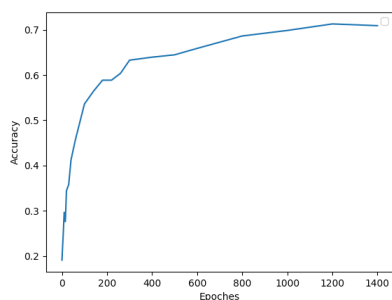


图 5.11 训练集大小与模型准确率之间的关系

图 5.11- (a) MLP 模型训练集大小与模型准确率之间的关系

图 5.11- (b) Spectral CNN 模型训练集大小与模型准确率之间的关系

图 5.11- (c) FirstCheb 模型训练集大小与模型准确率之间的关系

图 5.11- (d) DCNN 模型训练集大小与模型准确率之间的关系

图 5.11- (e) GCN 模型训练集大小与模型准确率之间的关系

图 5.11- (f) GAT 模型训练集大小与模型准确率之间的关系

图 5.11- (g) GraphSage-Mean 模型训练集大小与模型准确率之间的关系

图 5.11- (h) GraphSage-MeanPool 模型训练集大小与模型准确率之间的关系

图 5.11- (i) GraphSage-MaxPool 模型训练集大小与模型准确率之间的关系

从上图中我们可以看出，随着训练集大小的增加，模型的准确率会快速上升，随后趋近于稳定，训练集大小的增加对模型准确率的影响有所减小。通过观察发现，除了作为基线模型的 MLP 之外，Spectral CNN 模型表现最差，在训练集大

小在 50 和 200 之间时准确率增加过于平缓,而且在训练及大小在 600 到 1400 之间时未达到最大准确率,保持着一定速度的稳步增长。且 Spectral CNN 模型在训练集大小在 200 到 400 之间时不够稳定。DCNN 在测试的图神经网络模型中表现最不稳定,在测试集大小在 200 到 600 时测试集准确率出现了较为剧烈的波动,这种情况在其他模型上都没有出现。FirstCheb 和 Graphsage-Mean 表现最好,这两个模型在训练集大小为 200 时均基本达到最大准确率,训练集大小在 300 以后均已达到最大准确率。并且随着训练集大小的增长,这两个模型的测试集准确率均维持在最大准确率,没有出现波动或下降的状态。

但是,上述图像只表现了单个模型的测试准确率随训练集大小增长的状况,没有比较所有模型在训练集大小一定时准确率的差异。考虑到大部分模型在训练集大小超过 600 时已经达到最大准确率,并且模型准确率趋于稳定,受训练集增加而导致的准确率上升可以忽略不记。我们绘制在训练集大小在 0 到 600 之间的不同图神经网络模型预测准确率受训练集大小影响的折线图。为了方便对比,我们将所有模型的变化绘制到一个图像之中,得到的结果如图 4.12 所示:

通过图可以很直观的看出不同模型在相同的训练集大小下的准确度区别,从图中可以观测到,在训练集大小小于 40 时,Graphsage-Mean 模型的准确率较高,而当准确率大于 40 时,GCN 和 GAT 模型的准确率较高。对比上文中分析单个模型得到的结果,虽然 FirstCheb 模型的准确率随训练集大小增长而增长的速度最快。但是,FirstCheb 的准确率较低,虽然 FirstCheb 在较少的训练集上就可以发挥出模型的最大潜能,但通常并不是最好的选择。

综合测试数据,我们发现在训练集数量较少时(Citeseer 中训练集数量在 40 以下,即占总节点数量的 1.2%以下)时,选用 GraphSage-Mean 方法可以取得较好的效果。在训练集数量较多时,选择 GCN 和 GAT 模型可以取得较好的效果。

在图神经网络模型中,GraphSage-MaxPool 测试集准确率随测试集数量增长而增加的速度最为缓慢。在训练集数量较多时(Citeseer 中训练集数量在 600 以上,即占

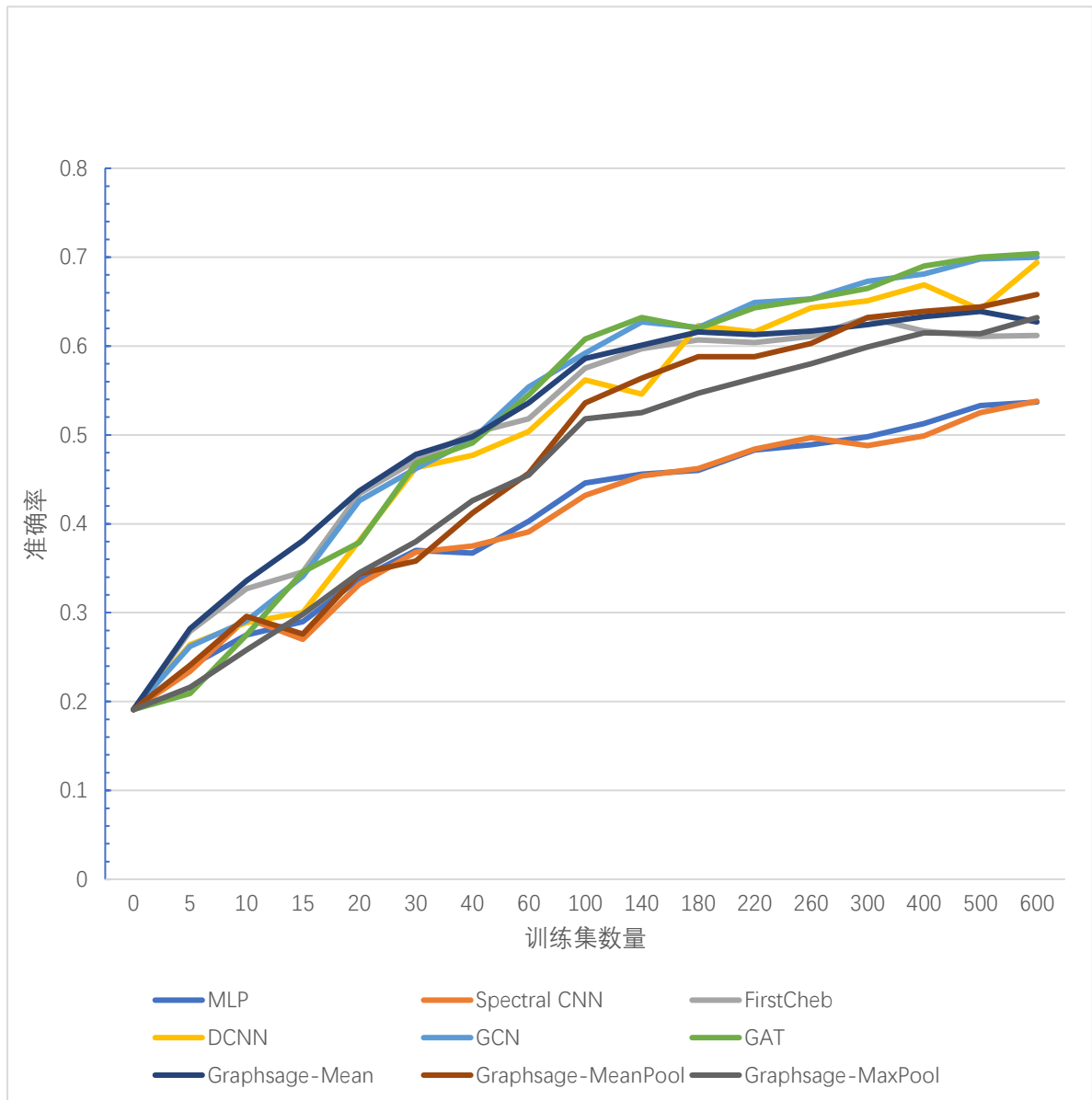


图 5.12 训练集大小-准确率折线图

总节点数量的 18) 时, 基本所有图神经网络模型均已基本达到最大准确率, 继续增加训练集大小不会对模型的准确率造成影响。