**Content**

**Basic:Reinforcement Learning**

1. **GraphNAS: Graph Neural Architecture Search with Reinforcement Learning**

**Basic:Biology**

1. **AffinityNet: Semi-supervised Few-shot Learning forDisease**

**Type Prediction**

**Title:**

**Basic:**

**Introduction:**

**Data:**

**Method:**

[**Evaluation**](javascript:;)**:**

**Advantage/Disadvantage:**

Advantage:

Disadvantage:

**Result:**

**Code:**

**Future Work:**

**Title:**

**GraphNAS: Graph Neural Architecture Search with Reinforcement Learning**

**Basic: NAS**

[x].Gao, Y., Yang, H., Zhang, P., Zhou, C., & Hu, Y. (2019). *GraphNAS: Graph Neural Architecture Search with Reinforcement Learning*. Arxiv.*April*.I <http://arxiv.org/abs/1904.09981>

[X].Gao, Y., Yang, H., Zhang, P., Zhou, C., & Hu, Y. (2020). ***Graph Neural Architecture Search*.** 1403–1409.ijcai.2020/195

**Introduction:**

**方案：**

**1.针对图神经网络设计搜索空间（Agent’s Action Space）**

**2.基于强化学习机制与LSTM设计搜索算法**

**实验验证：**

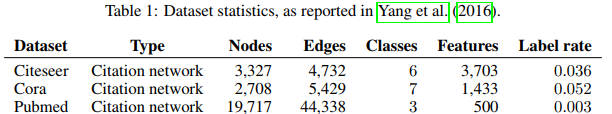
**1.半监督：学术文章分类**

**2.完全监督：蛋白质分类**

**Data:**

**1.Transductive Learning Task：**

**学术文献引用网络**



**数据集url：https://github.com/tkipf/gcn/tree/master/gcn/data**

**2.Inductive Learning Task：**

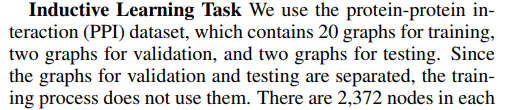
**PPI蛋白质相互作用网络**

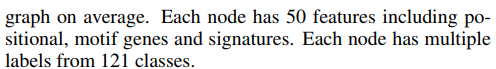
**(1).训练集：24 graphs,验证集：2 graphs, 测试集：2 graphs**

**(2).图平均节点数：2372**

**(3).单个节点特征维度：50**

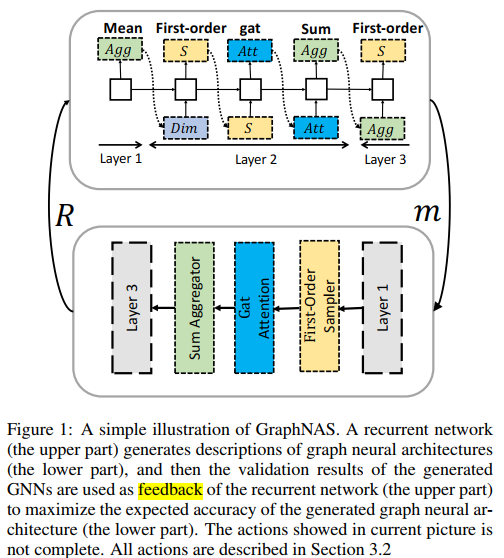
**(4).类别数：121类**



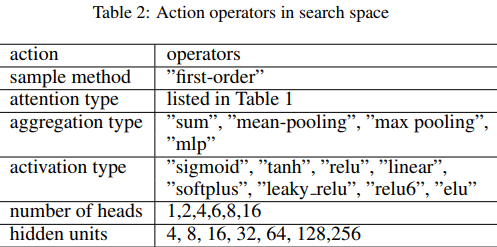


**Method:**

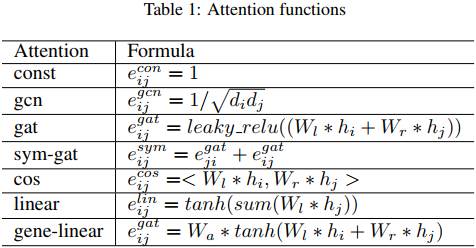
**模型架构：**



**1.搜索空间定义：**



**Attention方式：**



**2.基于强化学习机制的控制器设计:**

**2.1.单层LSTM**

**2.2.input\_dim：100(每个类型被编码为100维向量)**

**2.3.hidden\_dim：100**

**2.4.output\_dim: 100(有解码模块去解码出相应的action)**

**2.5.loss计算：以LSTM计算结果中1.熵信息、2.对数概率信息、3.LSTM设计出的GNN结构在验证集中验证的reward信息为基础，基于滑动平均的策略梯度算法确定loss值**

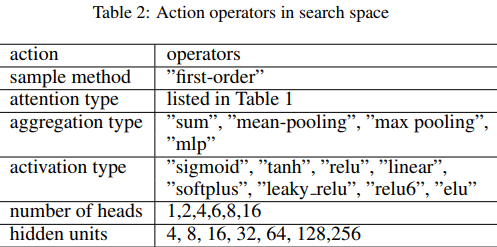
**2.6.optimization manner: adam**



**3.搜索过程：**

**3.1．搜索的GNN结构默认只有两层且全部使用一阶领域节点聚合信息**

**3.2．搜索顺序:每层的搜索顺序按下表从attention type开始以hidden units结束，其中number of heads是多头注意力机制中的头数。**



**4.Child GNN参数共享机制：**

**4.1．LSTM选出的GNN网络结构，网络结构参数基于体的操作action name用字典维护；**

**4.2.由LSTM产生新的GNN结构，结构中对应的参数从字典中获取为初始化参数**

**4.3.当新的GNN产生的reword为正时(在验证集上的精度高于历史最好精度)，相应的参数对字典中的参数进行更新(merge);**

**5.使用滑动平均处理多个Child GNN产生的reward得到最后的reward;**

**==================**

**6.对共享权重的探索：在固定LSTM参数下产生的新网络结构，重新训练共享参**

**数，重新积累reward修正LSTM参数**

**7.GNN结构最终结构确定：？？？**

**==================**

[**Evaluation**](javascript:;)**:**

**实验1：**

**AUC**

**试验2：**

**Micro-F1**

**Advantage/Disadvantage:**

**Advantage:**

**1.总结图卷积的有效架构并定义单层图卷积组件的搜索空间**

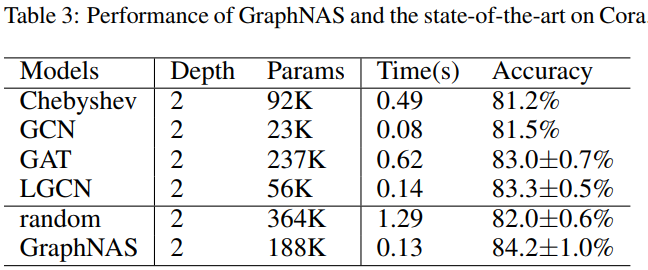
**2.使用强化学习进行搜索**

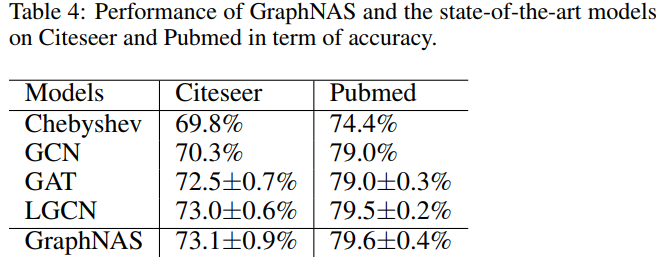
**Disadvantage:**

**1.搜索顺序是否可以更自由？**

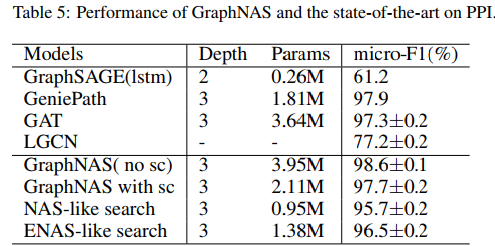
**Result:**

**实验1：**

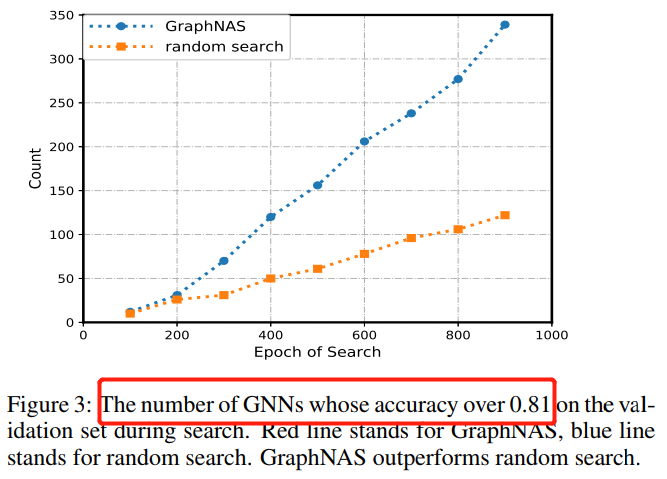


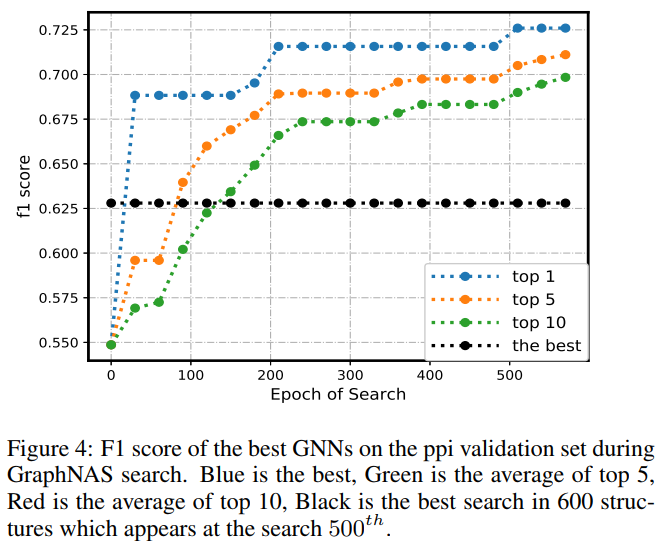


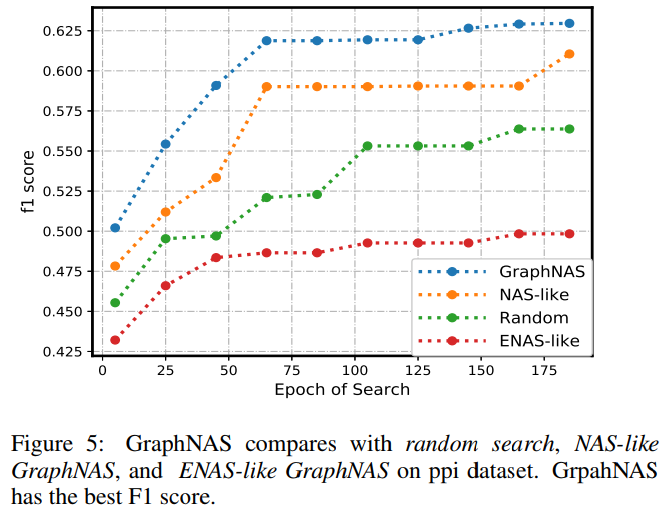
**实验2：**



**搜索效率与效果验证：**







**Code:可运行**

**Url:** [**https://github.com/GraphNAS/GraphNAS**](https://github.com/GraphNAS/GraphNAS)

**GNN库：**

**1.**[**https://github.com/dmlc/dgl**](https://github.com/dmlc/dgl)

**2.**[**https://github.com/rusty1s/pytorch\_geometric**](https://github.com/rusty1s/pytorch_geometric)

**Future Work:**

**Title:**

**AffinityNet: Semi-supervised Few-shot Learning forDisease Type Prediction**

**Basic:Biology**

[x].Ma, T., & Zhang, A. (2019). AffinityNet: Semi-Supervised Few-Shot Learning for Disease Type Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 1069–1076. https://doi.org/10.1609/aaai.v33i01.33011069

**Introduction:**

**方案：**

**1.基于attention相似度构图**

**2.基于特征提取器对单个特征进行特征筛选**

**3.基于GAT进行特征融合**

**4.小样本学习策略**

**实验验证：**

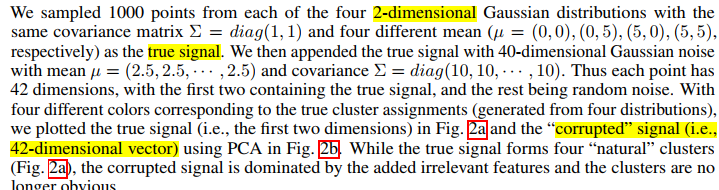
**1.癌症分类**

**2.癌症生存率估计**

**Data:**

**实验1：**

**42维特征向量=2维特征+40维高斯噪声，**

****

**数据集情况：**

**4个类，4000个样本**

**实验2:**

**url:** [**https://portal.gdc.cancer.gov**](https://portal.gdc.cancer.gov)

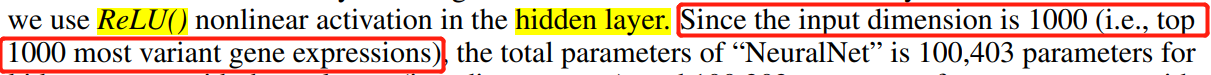
****

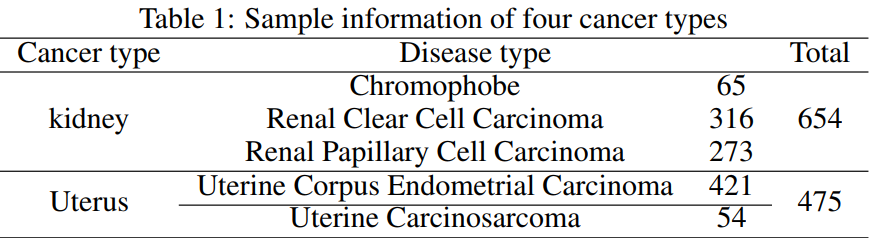
**数据集相关文献**

****

**数据集情况：**

**每个样本1000维**



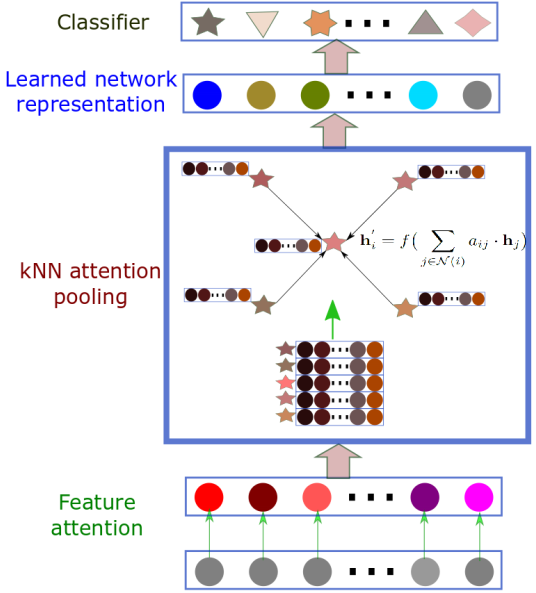


**实验3：**

**使用实验2中kindey数据集**

**Method:**

**模型架构:**

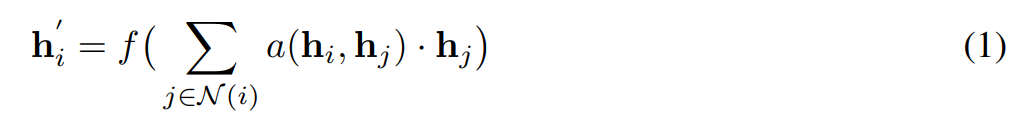


**1.Similarity graph 构造**

**基于attention核计算全部数据之间的相似度，以相似度作为构图依据构图**

**2.KNN attention pooling layer**

**(1).节点表达：**

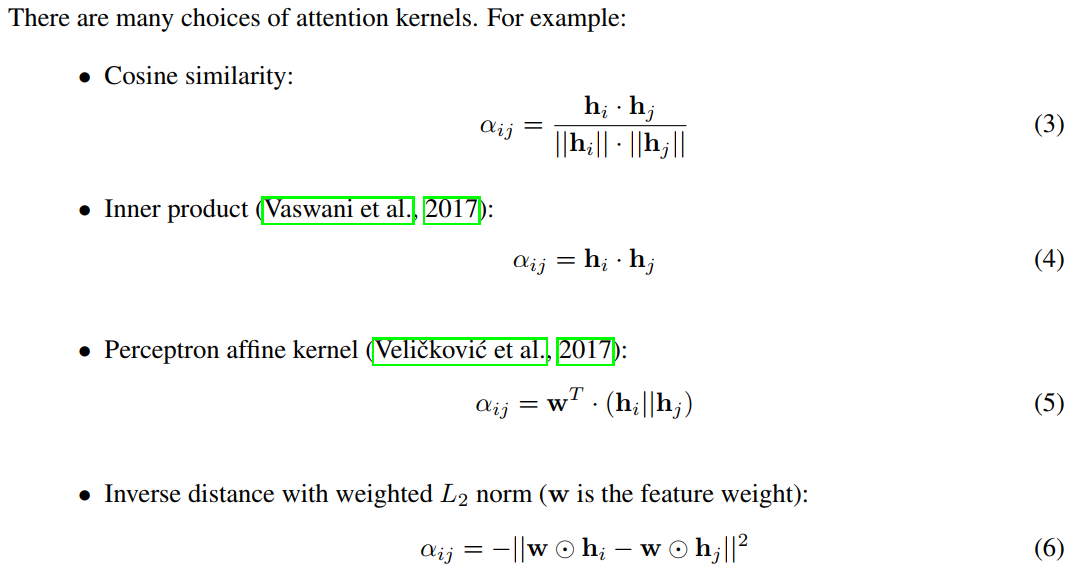


**(2).注意力机制：**



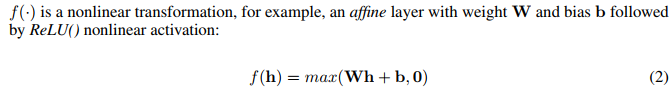








**(3).Relu激活函数：**



**(4).KNN Pooling操作：**

**对图规模大且节点具有高度的邻居节点选择，基于KNN思想，选择中心节点相似度最高的k个邻居节点聚合。**



**pooling在图像特征提取中的作用：**

**1.增加平移不变性**

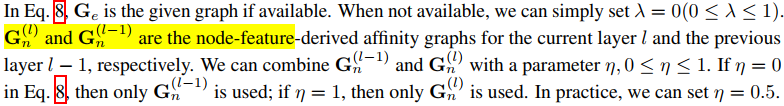
**2.保留主要特征/减少模型训练参数**

**本文pooling操作体现在基于attention kernel计算central node 与 neighbor nodes的similarity ,并以similarity来选择 k 个neighbor nodes起到减少central node聚合neighbor nodes数量的以达到类似图像pooling中第二个作用。**

**(5).Dynamic affinity graph操作：**

**每层图节点表示考了原始图节点表示，上一层图节点表示**





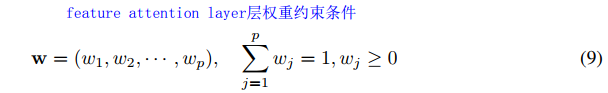
**4.Feature Attention Layer**

**通过有监督信号，对样本特征进行筛选**

**(1).节点表达**



**(2).权重约束**

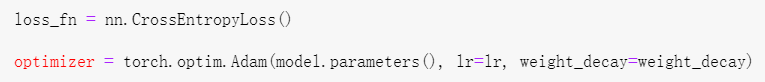


**5.Semi-supervised few-shot learning机制**

**6.损失函数与优化函数**

**损失函数：交叉熵损失**

**优化函数：Adam**



[**Evaluation**](javascript:;)**:**

**实验1：**

**1.training loss值**

**2.AUC**

**3.feature attention layer特征选择评估**

**实验2：**

**AMI评估**

**(https://www.jianshu.com/p/b9528df2f57a)**

**实验3：**

**1.Concordance index: 主要用于计算生存分析中的COX模型预测值与真实之间的区分度(https://www.jianshu.com/p/5e648f0f49ed)**

**2.基于Wilcoxon signed rank test的P值检验**

**3.Kaplan-Meier生存估计(https://zhuanlan.zhihu.com/p/97645982)**

**Advantage/Disadvantage:**

Advantage:

**1.数据组织：将独立非图结构数据依据相似度构图思想以图的形式进行组织并使用GAT对图数据进行融合表示;**

**2.卷积域选择：使用KNN思想选择卷积域范围并可实现每层卷积动态选择卷积域内邻居节点;**

**3.提出一种卷积层图表示融合的机制，增强卷积过程中图表示的平滑度;**

**4.加入单个样本特征选择机制并融入训练中学习选择权重;**

**5.基于训练好的特征选择器，定义聚类/分类计算中的相似度计算;**

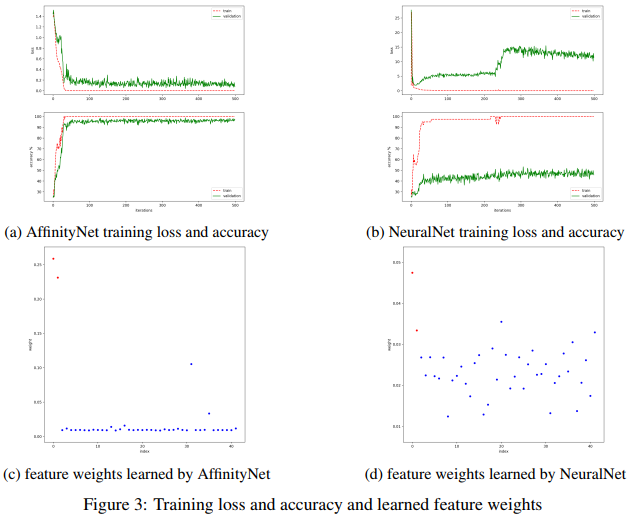
**6.做了很多生物信息学的实验来验证方法的有效性;**

Disadvantage:

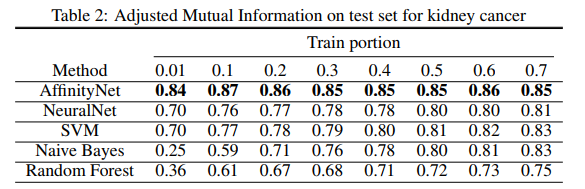
**1.对图结构数据挖掘没有特别优势**

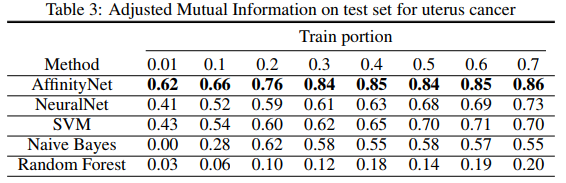
**Result:**

**实验1：**

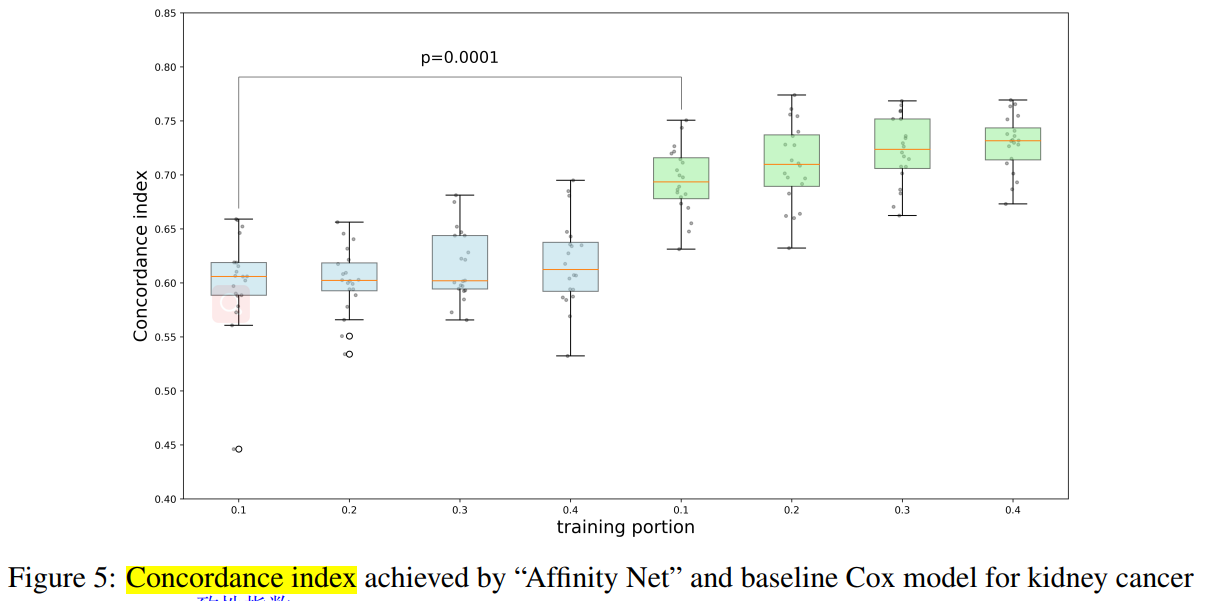


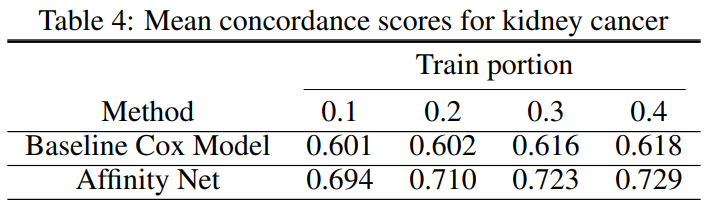
**实验2：**

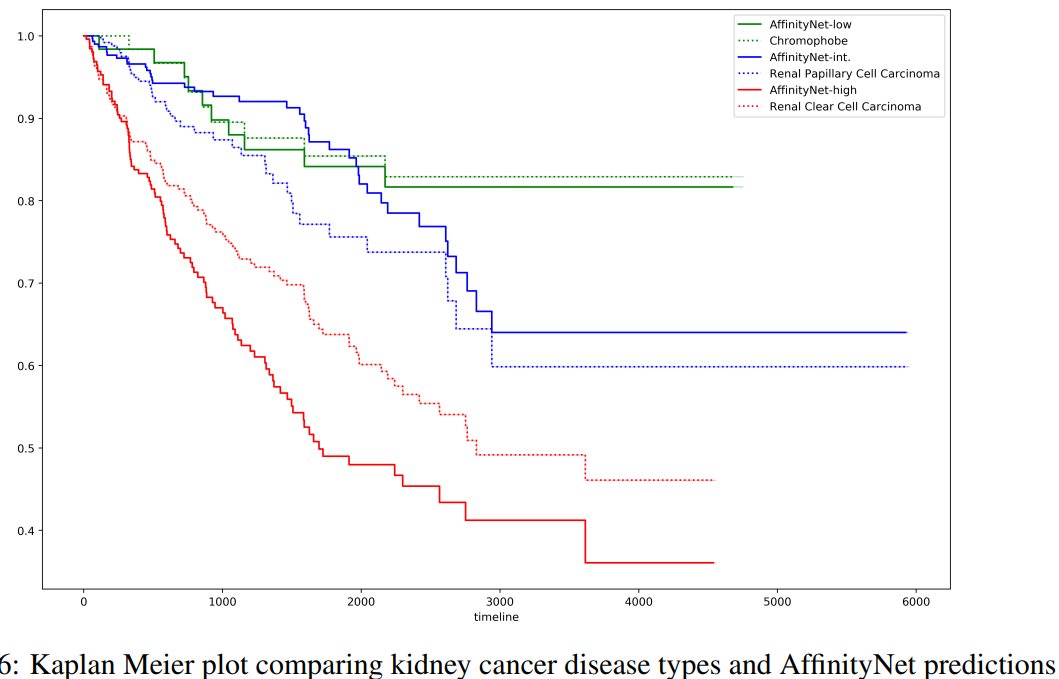




**试验3：**







**Code:可运行**

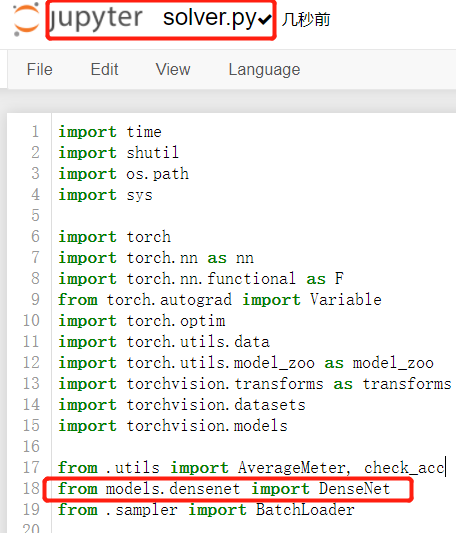
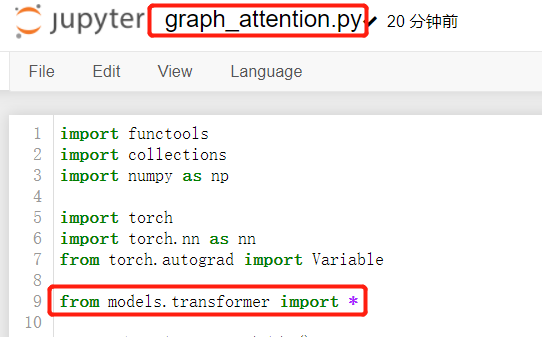
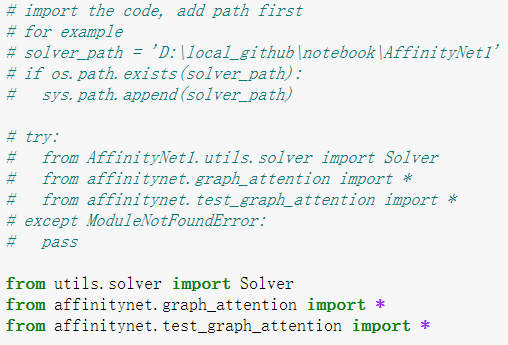
**url：https://github.com/BeautyOfWeb/AffinityNet.**

**复现实验1：**

**对比在不同比例训练情况下模型AUC与样本特征选择器效果**

****

**原始脚本修改：**



**Future Work:**