

QG 工作室周记

姓名：王伟杭

组别：数据挖掘组 年级：2020 级

周次：第 3 周

生活随记	这周状态回暖，去看了最近的怒火重案，谢霆锋好帅!!!
学习 开发 比赛 情况	<ol style="list-style-type: none">1.链路预测基于局部算法： CN、RA、AA 等2.链路预测基于随机游走（全局）算法： ACT、Cos+3.Word2Vec、随机游走、DeepWalk、Node2Vec4.链路预测指标 AUC、Precision
一周总结	<p>这一周做中期考核总体来说状态回暖很多，很多时候能够静下心来沉淀自己来学习算法，在图数据挖掘这个完全没接触过的领域学习了很多东西，读了一篇这个方面的入门文献以及跟着知乎和很多个人博客学到了很多这方面的算法。</p> <p>一二天的时间主要是在规划这七天的一个学习计划以及对链路预测这个方向做了一个比较细致的了解，对其的目标、方法做了一个初步的了解，结合路哥的 PPT 以及吕琳媛的复杂网络链路预测这篇文章，很快就接接下来该做的事情确定下来。接下来的几天就是着手于实现对节点的相似度进行评估，以及对算法得到的结果进行一个评估。这个部分一开始并不非常顺利，因为电脑的原因导致对这样一个大矩阵（10000+*10000+）数据无法进行运算，因此一开始都主要是在着手于降低算法的时间复杂度已经空间复杂度，后面芳琪实现了几个局部指标后，结合她的思路，很顺利的解决了这个问题，后续又花了一点时间对她实现的算法进行评估（AUC 和 Precision），遇到的一个很大的问题就是精确度很低，但是对算法的优化已经无从下手，遂转而学习基于随机游走的 ACT 以及 Cos+，这两个算法其实我并没有彻底理解他们的细节，只是跟着公式敲，遇到的一个问题就是又爆内存，只能把代码交给队友跑，有一点遗憾的是最终因为沟通的原因，没有对这两个指标进行评估，不知道这两个算法的实际实现效果如何。</p> <p>最后的几天，为了实现将社会属性考虑到预测中，将重心转向“如何利用社会属性”这个问题上，但是对这个问题始终没有找到比较有用的资料，只能够漫无目的的探索。为了解决对含权网络的预测，转而学习更加先进的 Embedding 方法：Node2Vec，作为一个 Deepwalk 的改进，抱着循序渐进的态度，又先学习了 Deepwalk 的前置知识：随机游走以及 Word2Vec。可以说我最后的时间都是花在这上面，尤其是对 Word2Vec 的学习，</p>

	算是比较系统的对其细节了解了一遍。比较可惜的是找到了两个这方面的库，但是由于python 基础的问题，没能应用到系统上，导致了最终系统使用的算法还是比较基础的AA。
存在问题 未来规划	1.对 embedding 的理解不够透彻，花了很多时间在学习嵌入法； 2.数学基础不足，必须在有空的时间加强学习数学，才能更好的理解机器学习的算法 3.与队友的沟通对接不足，没有及时获取队友的反馈 4.要多沉淀！多反思！要多沉淀！多反思！要多沉淀！多反思！ 5.对 python 的一些功能以及规范还需要继续进行学习。
导师评价	