大创问题

1. 现在为项目做了哪些准备，目前的进度到哪了？

答：我们目前在第一阶段，也就是知识准备工作。主要学习基础分析方法、机器学习相关的统计学习模型与算法知识、了解掌握智能运维体系的基本架构以及学习Linux Shell脚本语言。其中有些知识在专业课上会涉及到，其余的我们将利用课余时间自学。

2. 经费怎么用你们有规划吗？

答：基本上就是服务器和学习资料购买,以及论文的版面费。

3. 如何收集数据？

答：我们的导师会帮助我们收集数据，收集企业的基站KPI数据。

4. 原理上你们了解多少？

答：原理上就是在原来的深度学习的异常检测算法上结合聚类的算法。

5. 最后预期的成果产出是什么？

答：我们的结果是，训练出一个可以分析基站收集的多维KPI信息，来发现、检测和定位设备异常的无监督多指标异常检测模型。该模型最终能及时反馈无线网络出现的问题，从而保障网络连接畅通。同时我们预期产出一篇论文。

1. 你们对模型评价指标有了解吗？

答：目前我们预期对模型处理后的指标设定阈值从而平衡模型的性能，具体情况需要按照模型设计后来做出抉择。

7. 具体的聚类实现思路：

解决方案 （1）聚类：OmniCluster，计算两两数据之间的距离（欧氏距离，SBD距离），层次聚类（HAC）不受距离极端值的影响，按照距离的排序来聚类，比较鲁棒。 （2）将多指标异常检测模型和聚类结果结合起来，一个簇一个模型。簇的数量是远远小于基站（曲线）的数量的，大大降低了模型训练的开销。



