任务需求

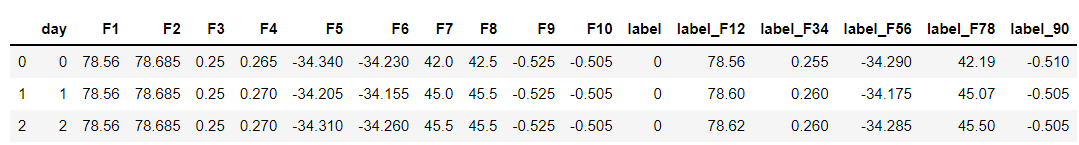
1. 简单地说，这个任务的目的是预测，但分为了两个部分：**预测故障状态**（分类任务）和**预测性能指标**（也就是回归任务）。该任务有两个目的：（1）复习我们学过的深度学习相关知识，能够灵活运用神经网络（更多的知识需要自己学习）解决分类和回归问题；（2）实验室提供的数据集与项目中的数据类似，该任务能够检验为项目提出的解决方案。
2. 数据介绍：
   1. 数据说明：该数据是光网络的单板性能数据，共**775**块单板采集周期为**47**天。
   2. 数据排列顺序为每1块单板的0-46天数据、第2块单板的0-46天数据、…、第775块单板的0-46天数据。
   3. 数据集的大小为**36425** 行× **17** 列。
   4. 数据特征及标签说明：

图1、数据节选

表1、数据集的特征及标签对应含义的说明

|  |  |  |
| --- | --- | --- |
|  | 特征或标签 | 含义 |
| 1 | day | 每一块单板数据所代表的天数，用以表示顺序。 |
| 2 | F1 | 激光器偏置电流最小值(mA) |
| 3 | F2 | 激光器偏置电流最大值 |
| 4 | F3 | 输出光功率最小值(dBm) |
| 5 | F4 | 输出光功率最大值 |
| 6 | F5 | 输入光功率最小值(dBm) |
| 7 | F6 | 输入光功率最大值 |
| 8 | F7 | 单板环境温度值最小值(℃) |
| 9 | F8 | 单板环境温度值最大值 |
| 10 | F9 | 激光器温度偏移值最小值(℃) |
| 11 | F10 | 激光器温度偏移值最大值 |
| 12 | label | 关于该单板在这一天是否发生故障（0为正常，1为故障） |
| 13 | label\_F12 | 激光器偏置电流平均值（F12代表F1和F2所对应的标签，如下同理） |
| 14 | label\_F34 | 输出光功率平均值 |
| 15 | label\_F56 | 输入光功率平均值 |
| 16 | label\_F78 | 单板环境温度值平均值 |
| 17 | label\_F90 | 激光器温度偏移值平均值 |

说明：第1行用于表征单板和时间的不同；第2-11行为单板性能数据特征；第12行为分类标签；第13-17行为回归任务的标签。分别对应着不同的颜色。

1. 任务目标

总的任务就是根据所学进行神经网络模型的搭建，通过训练使模型分别能够完成预测是否故障和预测性能指标的两个任务。并且在测试集上获得更低的**误报率**和**漏报率**，或者是得到更大的预测范围（**即N越大**）。

* 1. 不论分类还是回归，我们的目标都是预测，也就是通过前**X（X>=1）**天的数据推测后**N（N>=1）**天的数据。因此，第一步需要对原始数据集进行重新构建，获得前X天（特征）与后N天（标签）的对应关系。
  2. 搭建合适的网络，根据任务需要选择**神经网络**或**拼接所需的层**来提高预测效果。
  3. 根据任务需要完成代码的编写，包括：损失函数，优化器，评价指标等。

1. 具体需求
   1. 数据集重构阶段需要探究的内容包括但不限于：是否需要标准化；**X和N参数**的选择，数据是否平衡（解决数据不均衡具体实现）等。
   2. 网络搭建阶段需要研究的内容包括但不限于：该选择哪个模型作为baseline，能有那些改进如是否加BatchNorm，Dropout，以及中间层维度的选择。即对各种明参进行调参（可以搜集一些调参的资料进行学习，如网格搜索，随机搜索等）。
   3. 网络训练要考虑的内容包括但不限于：对于一些明参进行调节，如学习率，batch，epoch的选择，探究如何确定模型训练可以结束等。
   4. 预测回归任务类似于项目中的解决方案，可以多查阅资料与项目进行关联。
2. 任务提交及总结
   1. 我会在GitHub上新开一个repository，代码和文档写一部分就可以提交一部分，**需要写注释**，不需要定期进行提交。最后截止日期为开学时。**Ps：请不要在库中上传原始数据集。**
   2. 每一个部分的训练选择需要**详细**说明，用**数据**和**图表**进行阐释，画图请使用**Origin画图软件**，不会的可以从B站上学习。内容格式可以按照附件中的**模型文档**进行描述。
   3. 写代码过程中遇到的**报错**或者**Bug**，请登记在**问题解决文档**中，解决后，写明错误原因，如何解决，并附上Reference。供大家分享纠错。