HAPTS源码说明

# 进度说明

## 目前进度

本项目通过对UCI 上的数据集Smartphone-Based Recognition of Human Activities and Postural Transitions Data Set（HAPT）数据集作为试验数据集，选用HAPT Data set与RawData数据完成了对其数据进行处理，并搭建模型训练、预测。

## 已完成的内容

### 内容一

本次实验采用的是RawData以及HAPT Data set的数据，其中HAPT Data set的数据是包含561列的特征数据，本实验中对数据进行了归一化处理。而RawData数据是以每一个实验为一个文件的数据组，本实验中对其进行了整合为了一个csv文件，保留了8列数据（x，label，grox，groy，groz,accx，accy，accz）。并实现了归一化。

### 内容二

对HAPT Data set文件夹的数据集对其使用bp神经网络进行处理，得到了94%的准确率。对RawData的数据使用了滑动窗口法进行了处理：每次取256组两传感器三轴数据，并取50%重叠；使用了cnn与lstm模型进行训练，得到了90%准确率。

### 内容三

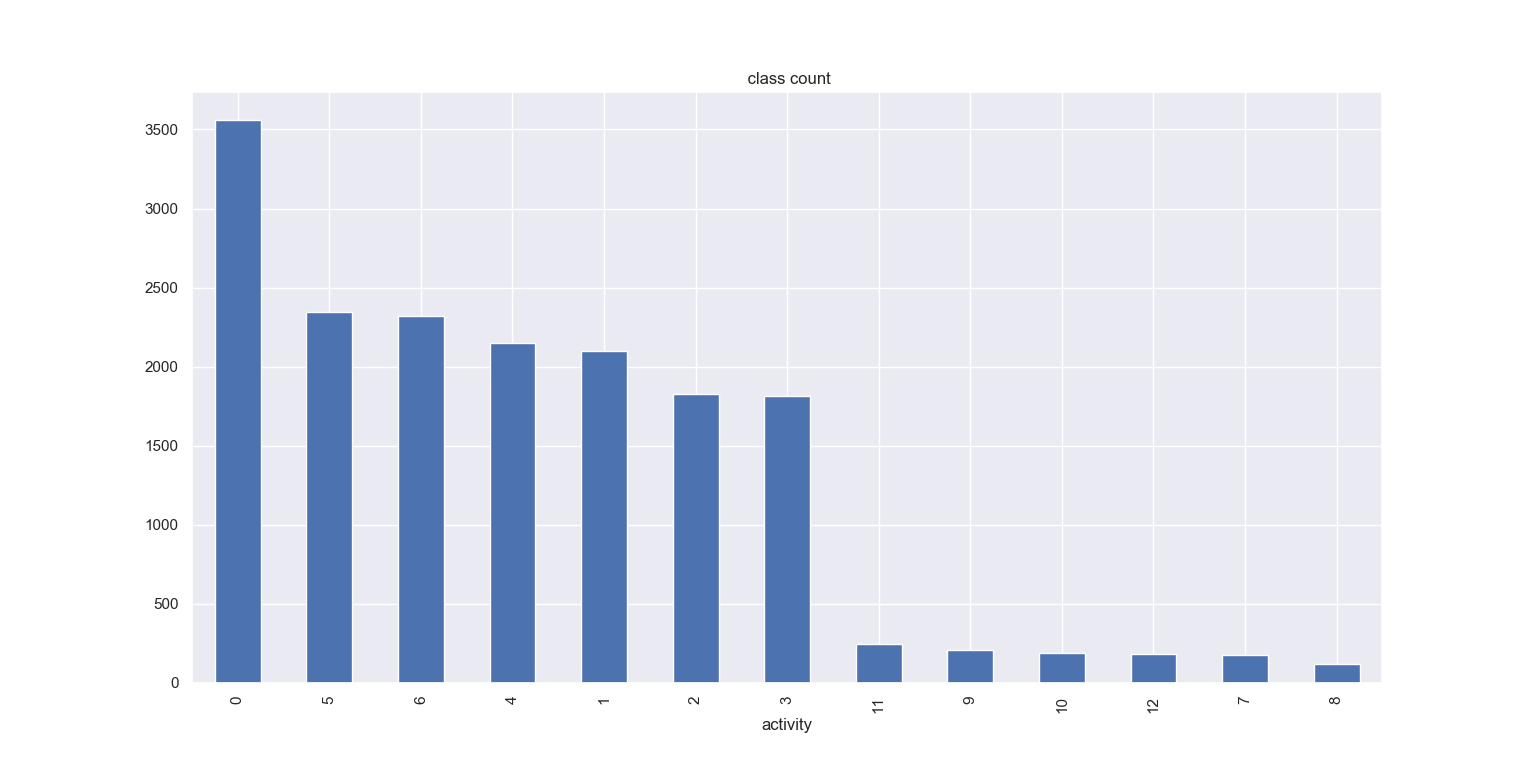
在上述的TensorFlow代码基础上，学习飞桨模型，将代码转化成paddle代码，在飞桨上实现了bp神经网络与lstm，得到了近似的准确率。

## 需要改进的地方（功能欠缺的部分）

### 1.3.1改进一

数据类别不平衡，如下图所示，数据的类别出现了较大的倾斜，7/8/9、10/11/12类(STAND\_TO\_SIT，SIT\_TO\_STAND，SIT\_TO\_LIE ，LIE\_TO\_SIT

，STAND\_TO\_LIE，LIE\_TO\_STAND)的数量较少，且未采用过采样算法如smote等使得数据类型平衡；因此准确率无法明显提升。



### 1.3.2 改进点二

由内容二可知，使用简单的bp神经网络可以在561维的数据上取得了较高的准确率，未实现将RawData数据提取为561维特征向量的数据。

## 1.3.3 改进点三

网络结构的调整，在RawData数据集上使用了滑动窗口法与cnn，lstm模型训练预测，但未对网络结构以及参数做出过多的调整，使得准确率没有得到更多的提高。对于paddle上的lstm的模型的学习较少，导致在paddle的准确率较低。

# 重难点说明（便于审核人员理解）

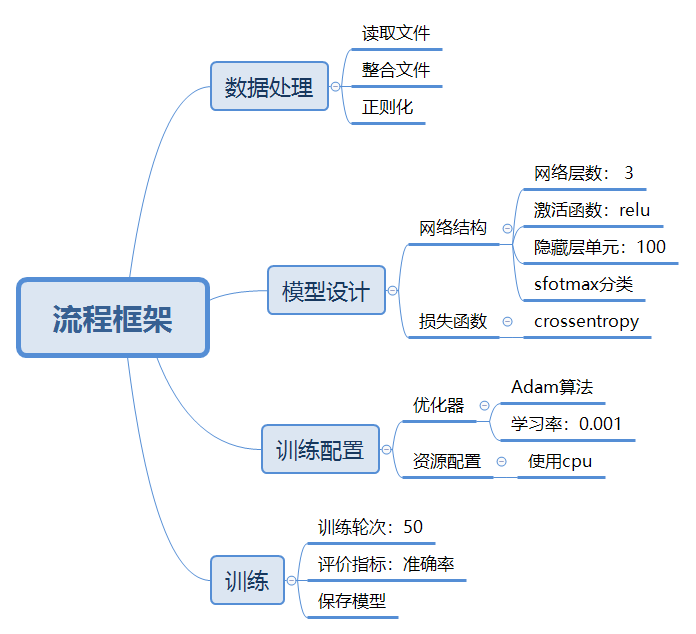
## 重难点一

在将tensorflow的代码修改为paddle代码时遇到了问题。如variable用法与fluid中create\_param的用法有些许不同，在数据读入时，paddle上是无需进行独热编码操作，且进行了独热编码操作后得到的错误后在网上找不到对应的；在paddle上出现了没有对应的接口，如tf.conv1d，通过浏览参考百度ai studio的项目总结解决了问题。

## 重难点二

难点说明

在调整网络结构、参数上花费了较多的精力。



如上图是用飞桨实现的bp网络的框架，以这个bp网络为例子，可调整的有网络层数，学习率，激活函数类型，lstm的框架同理；此外，对于raw data数据集上，对于滑动窗口取的数据多少，重叠率为多少也需要考虑。由于对于网络结构以及参数调整的经验过少，每次调整、训练花费的时间较多。

# 源码版本更替说明

## 版本一

1. 实现了数据处理
2. 基于tensorflow上实现了bp神经网络对数据进行训练。

## 版本二

添加了运用滑动窗口法对RawData传感器数据进行处理，并基于tensorflow实现了lstm与cnn模型对数据训练，预测。

## 版本三

添加了基于tensorflow修改的paddle代码，实现了lstm，bp神经网络模型对数据进行训练，调整了网络结构以及参数。