**吊机故障预测的隐马尔科夫模型使用说明**

1. 模型用途

给定吊机历史数据，根据吊机历史数据来预测下一时间节点设备是否会发生故障，有助于及时发现并排查问题。

1. 数据说明

输入数据用变量x表示，共有8个特征：

第一维元素x(0)表示：使用时间 (h),

第二维元素x(1)表示：吊机长度 (m),

第三维元素x(2)表示：每吊速度 (m/min),

第四维元素x(3)表示：吊机横截面积 (cm^2),

第五维元素x(4)表示：每吊重量 (kg),

第六维元素x(5)表示：吊数 (可用1、2、3、4表示),

第七维元素x(6)表示：吊机操作人员的人数,

第八维元素x(7)表示：吊机人员的熟练度 (一个0-1之间的数，越大表明这个团队工作效率越高)。

输出数据用y表示，只有1个特征：

是否会发生故障。

1. 算法原理

吊机故障预测隐马尔科夫模型的核心思想是将设备的历史数据看作为一个可观测的状态集合，每一份历史数据[x(x(0),x(1),…,x(n),n为特征数),y]都是集合中的一个可观测状态，而是否发生故障可以看作为一个不可观测的隐藏状态，

在上面的这些情况下，可以观察到的**状态序列和隐藏的状态序列是概率相关**的。于是我们可以将这种类型的过程建模为有一个隐藏的马尔科夫过程和一个与这个隐藏马尔科夫过程概率相关的并且可以观察到的状态集合，就是**隐马尔科夫模型**。

（一）数据预处理

吊机故障预测隐马尔科夫模型的输入数据共有八个维度，然而各个特征之间量级有很大的差异，所以其对于数据的数量级十分敏感，数量级的过大差异将导致量级较大的特征占据主导地位。所以我们需要对与这些数据特征进行标准化，将它们都缩放到相似的范围，同时标准化还将加快模型训练时的迭代速度。

吊机故障预测隐马尔科夫模型可以使用两种方法对各个特征进行数据的标准化：min-max标准化和z-score标准化。其中，min-max标准化的公式为:, min和max分别为第i个特征的最小值和最大值。这可以将原始值映射到[0,1]区间内。z-score标准化的公式为,和分别为第i个特征对应样本集的均值和标准差。这使得原始值满足正态分布。

（二）数学建模

HMM主要解决的三个问题。

假设隐藏状态序列和观测状态序列分别使用Z和X表示，则解决的3个问题可表示为:

1.解码问题：已知模型参数和X，估计最可能的Z；维特比算法

2.概率问题：已知模型参数和X，估计X出现的概率；向前-向后算法

3.学习问题：仅给出X和隐藏层个数，估计模型参数。 B-W算法，通常是经过一定数量的训练以后，得到模型，然后解决问题1和2。

1. 使用方法

运行train.py文件依据历史数据得到模型参数（仅用于训练模型，正常使用时不需运行），具体输入命令：

**python train.py --train\_dir 训练数据存放地址**

例如：

**python train.py --train\_dir ./trainDataHJ.txt**

之后根据模型参数，运行model.py文件得到预测工时，具体输入命令：

**python model.py --train\_dir 训练数据存放地址 --predict\_data 预测数据存放地址**

例如：

**python model.py --train\_dir ./trainDataHJ.txt --predict\_data ./predictData.txt**

预测结果会存放在**./result/predictResult.txt**文件中