**零件产品质量问题与环境的关联分析模型**

1. 模型用途

给定产品的各类环境因素使用情况，通过零件产品质量问题与环境因素的相关历史数据，分析零件产品质量问题受到哪些环境因素的影响。

1. 数据说明

输入数据用变量表示，共有个特征，表示参与操作过程中各类人员的历史工作情况，为保持模型的灵活性人员个数可由参数控制：

第一个特征表示：第一类环境因素,

第二个特征表示：第二类环境因素,

第三个特征表示：…,

.

.

.

第个特征表示：第m类环境因素。

输出数据用y表示，只有1个特征：

本轮作业的零件产品质量问题。

1. 算法原理

零件产品质量问题与环境因素的关联分析模型的核心思想是将每类环境因素变化的历史序列看做是二维空间的一条曲线，我们将其称为子序列；同时也将本轮作业的零件产品质量问题也看作是二维空间的一条目标曲线，我们将其称为母序列。然后分别计算目标曲线和每一条环境因素曲线的关联程度。所谓关联程度，实质上是曲线间几何形状的差别程度。两条曲线间的几何形状差别程度越小，则关联程度越大，反正则关联程度越小。最后按母序列与各个子序列之间的关联程度大小排序返回关联程度最大的前个子序列。

（一）数据预处理

零件产品质量问题与环境因素的关联分析模型的输入数据所代表的每类人员的历史工作用时的各个特征之间量级可能由于工作特点不同而有很大的差异，例如操作路段，配件种类等特点的不同。因此，模型需要首先对输入数据进行预处理。

进行零件产品质量问题与环境因素的关联分析时模型可以使用两种方法对各个特征进行无量纲化处理：

（1）初值化处理:

（2）均值化处理：

（二）数学建模

（a）计算关联系数

所谓关联程度，实质上是曲线间几何形状的差别程度。因此曲线间差值大小，可作为关联程度的衡量尺度。对于一个参考数列 ，有若干个比较数列 ,各比较数列与参考数列在各个时刻（即曲线中的各点）的关联系数可由下列公式算出：

其中

（b）关联度排序

因为关联系数是比较数列与参考数列在各个时刻（即曲线中的各点）的关联程度值，所以它的数不止一个，而信息过于分散不便于进行整体性比较。因此有必要将各个时刻（即曲线中的各点）的关联系数集中为一个值，即求其平均值，作为比较数列与参考数列间关联程度的数量表示，关联度公式如下：

值越接近1，说明关联性越好。

因素间的关联程度，主要是用关联度的大小次序描述，而不仅是关联度的大小。将m个子序列对同一母序列的关联度按大小顺序排列起来，便组成了关联序，它反映了对于母序列来说各子序列的“优劣”关系。最后，可以根据关联序找到与零件产品质量问题母序列最相关的几个子序列对应的人员。

1. 使用方法

运行model.py文件得到与零件产品质量问题关联性最强的两类环境因素，具体输入命令：

**python model.py -i <inputfile> -o <outputfile>**

其中“i”参数后是零件产品质量问题与环境因素数据存放文件名 ，“o”参数后返回关联分析结果。

例如：

**python model.py -i ./** **productDataEnv.txt –o ./RAResult.txt**

模型结果会存放在**./RAResult.txt**文件中