HUAWEI

7.1 机器学习助力预测性维护(下)

目录

Contents

预测性维护难点

设备的预测性维护的难点在于故障数据相对正常运行来说是非常少的，并且对于设备的监测往往是传感器按时间采集的震动、位移等等信息，信息量较少。

传感器采集的信息是时序信息，且在时间长度上存在着长短不一的情况，在做设备的预测性维护之前，需要将传感器采集的原始数据进行一系列的变换，进行特征的生成，将基于时间的时序信息转换为非时序信息。

特征处理与变换是预测性维护当中非常重要的第一步。

目录

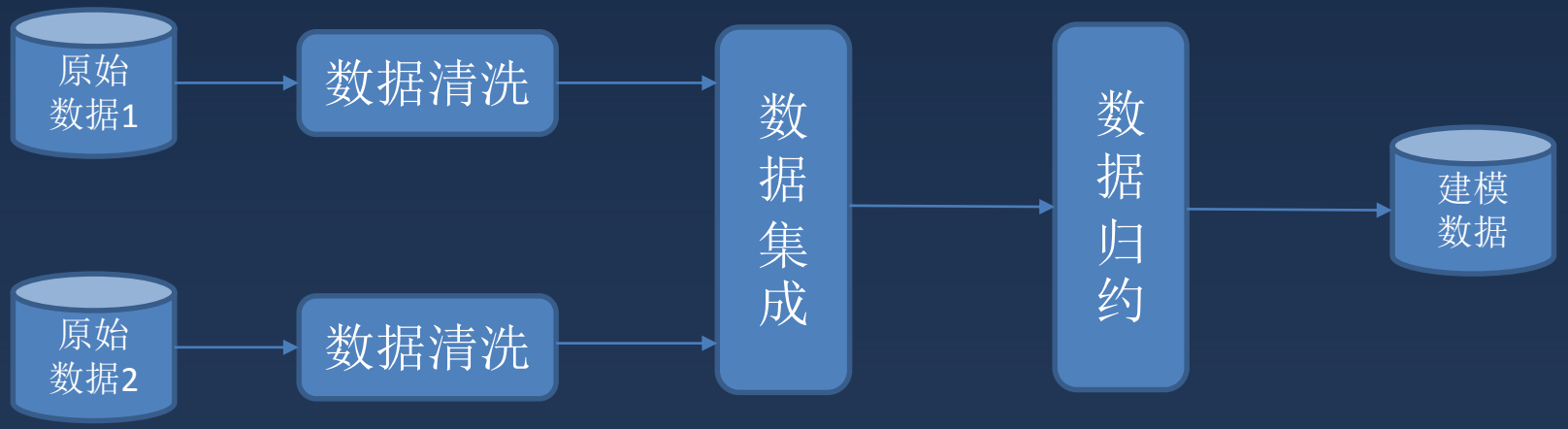
Contents

数据预处理的意义

机器学习建模需要高质量的数据：准确、完整、一致、时效、可信。

实际应用场景中的数据存在：不完整、不正确、含噪声。

一句略显夸张的话：数据处理3个月，建模5分钟



特征生成

特征生成

组合：多个特征按照一定规则形成新的特征，比如加和、相乘等

特征统计：在某个特征上采用一些统计量或者特殊的计算规则生成新的特征

特征移位：在时间相关的一些场景，可以将特征进行移位操作，比如周1的A特征和周2的B特征组合成为新的特征。

特征组合举例

原始特征<x_1，x_2，x_3>，组合特征<x_1x_2，x_2x_3，x_1x_3，x_1^2，x_2^2，x_3^2>

特征统计举例

原始特征<x_1>，将样本分片，统计特征<mean(x_1)，max(x_1)，min(x_1)，variance(x_1)，median(x_1)，...>

特征移位举例

原始数据

x1	x2	x3
1	0.11	123
2	0.23	456
3	0.45	789
4	0.67	1001

特征移位后数据

x1	x2	x3_new
1	0.11	456
2	0.23	789
3	0.45	1001
4	0.67	null

目录

Contents

特征选择

特征选择：面对成百上千维的特征，大概率存在着与目标工作无关的特征，是冗余的，所以需要一些方法进行特征选择。

主成分分析：将n维特征经过principal compos analysis或PCA搜索k个最能代表数据信息的维度，在搜索、计算的过程中维度本身发生了改变，产生的是原始维度的映射，将映射后的特征进行重要性排序。

特征子集：当存在n个原始特征时，即有2^n个特征子集。在n个特征进行筛选时，可以使用统计显著性检验的方法、信息增益的方法、决策树的方法。

相关性分析：用于检验2个特征之间的相关性，主要指标有Pearson相关系数、Spearman相关系数、卡方检验等。

主成分分析步骤：

1、在每个特征上减去这个特征的平均值。2、计算数据集的协方差矩阵。3、计算协方差矩阵的特征值和特征向量。4、将特征值排序保留最大的N个特征值。5、将数据转换到特征向量构建的新空间中。

特征子集主要方法：

1、逐步向前，从空集开始，每一次添加当前特征集中最有价值的特征，特征价值来自于统计显著性检验、信息增益等。

2、逐步向后，从特征全集开始，每一次删除当前特征集中最没有价值的特征，特征价值来自于统计显著性检验、信息增益等。

3、决策树，构建一棵规定深度的决策树，将没有出现在树中的特征删除。

特征归约

特征归约：将特征划分到类似的空间当中，消除彼此度量差异造成的影响。

离散化：将连续型特征映射到一个类别当中，比如收入的具体数值映射为“高、中、低”

标准化：将所有的连续型特征都归约到同一个区间上。常用的区间是[-1，1]和[0，1]

光滑：与特征清洗类似，采用回归等技术去除特征上的噪声数据。

聚集：不关注每一条样本的特征信息，关注样本集分片后的特征信息，等同于在特征上分片。比如原始样本是日销售额，而建模可以只关注周销售额。

离散化方法：

1、等宽分箱（连续值的区间划分是宽度一致的）
2、等频分箱（连续值的划分是出现次数一致的）
3、聚类分箱（按照特征上聚类的结果进行分箱）
4、邻近分箱（在分类问题中选择跟自己最近的类别标签相同的值进行合并）

标准化方法：

1、最大最小值标准化
2、Z-score标准化
3、和值比例标准化

数据说明

有1个故障监测时序文件，有两列<rawdata,output>，表示两种维度上的测量。这个时序文件表示在这个时序持续期内，该设备是处于故障”Abnormal1”状态的。

数据片段

	rawdata	output
0	-8.91235	1326.09
1	-9.33105	907.492
2	-9.57031	668.334
3	-9.98901	249.731
4	-10.1685	70.3846
5	-10.647	-408.035
6	-10.946	-707.012

代码实现

详见操作指导

Thank You.

华为云机器学习服务MLS
www.huaweicloud.com/product/mls.html