

# 家用电器用户行为分析与事件识别

- 1) 根据热水器采集到的数据, 划分一次完整用水事件。
- 2) 在划分好的一次完整用水事件中, 识别出洗浴事件。

## 分析方法与过程

- 1) 对热水用户的历史用水数据进行选择性抽取, 构建专家样本。
- 2) 对步骤 1) 形成的数据集进行数据探索分析与预处理, 包括探索用水事件时间间隔的分布、规约冗余属性、识别用水数据的缺失值, 并对缺失值进行处理, 根据建模的需要进行属性构造等。根据以上处理, 对用水样本数据建立用水事件时间间隔识别模型和划分一次完整的用水事件模型, 再在一次完整用水事件划分结果的基础上, 剔除短暂用水事件, 缩小识别范围。
- 3) 在步骤 2) 得到的建模样本数据基础上, 建立洗浴事件识别模型, 对洗浴事件识别模型进行模型分析评价。
- 4) 对步骤 3) 形成的模型结果应用并对洗浴事件划分进行优化。
- 5) 调用洗浴事件识别模型, 对实时监控的热水器流水数据进行洗浴事件自动识别。

## 数据抽取

### 数据探索分析

用水停顿时间间隔为一条水流量不为 0 的流水记录同下一条水流量不为 0 的流水记录之间的时间间隔。

### 数据预处理

#### 数据规约

- 属性规约
- 数值规约

#### 数据变换

目标是对用户的洗浴事件进行识别, 这就需要从原始数据中识别出哪些状态记录是一个完整的用水事件, 从而识别出用水时间中的洗浴事件; 一次完整的用水事件是根据水流量和停顿时间间隔的阈值取划分, 因此简历阈值寻优模型。为了提高在大量的一次完整用水事件中寻找洗浴事件的效率, 简历筛选规则剔除可以明显判定不是洗浴的事件, 得到建模数据样本集。

- 一次完整用水事件的划分

- 1) 读取数据记录, 识别到第一条水流量不为 0 的数据记录记为  $R_1$ , 按顺序识别接下来的一条水流量不为 0 数据记录为  $R_2$ 。
- 2) 若  $gap_i > T$ , 则  $R_{i+1}$  与  $R_i$  及之间的数据记录不能划分到同一次用水事件。同时将  $R_{i+1}$  记录作为新的读取数据记录的开始, 返回步骤 1); 若  $gap_i < T$ , 则将  $R_{i+1}$  与  $R_i$  之间数据记录的划分到同一次用水事件, 并记录接下来的水流量不为 0 数据记录为  $R_{i+2}$ 。
- 3) 循环执行步骤 2), 直到数据记录读取完毕, 结束事件划分。

- 用水事件阈值寻优

固定的停顿时长阈值对于某些特殊的情况的处理是不理想的, 存在把一个事件划分为两个事件或者两个事件合为以一个事件的情况。

当存在一个阈值的斜率指标  $K < 1$  时, 则取阈值最小的点  $A$  (可能存在多个阈值的斜率指标小于 1) 的横坐标  $x_A$  作为用水事件划分的阈值, 其中  $K < 1$  中的 1 是经过实际数据验证的一个专家阈值。

当不存在  $K < 1$  时, 则找所有阈值中斜率指标最小的阈值。

- 属性构造

时长、频率、用水的量化和用水的波动。

- 筛选候选洗浴事件

- 数据清洗

## 模型构建

模型检验