

LSM6DSV16X：机器学习内核

简介

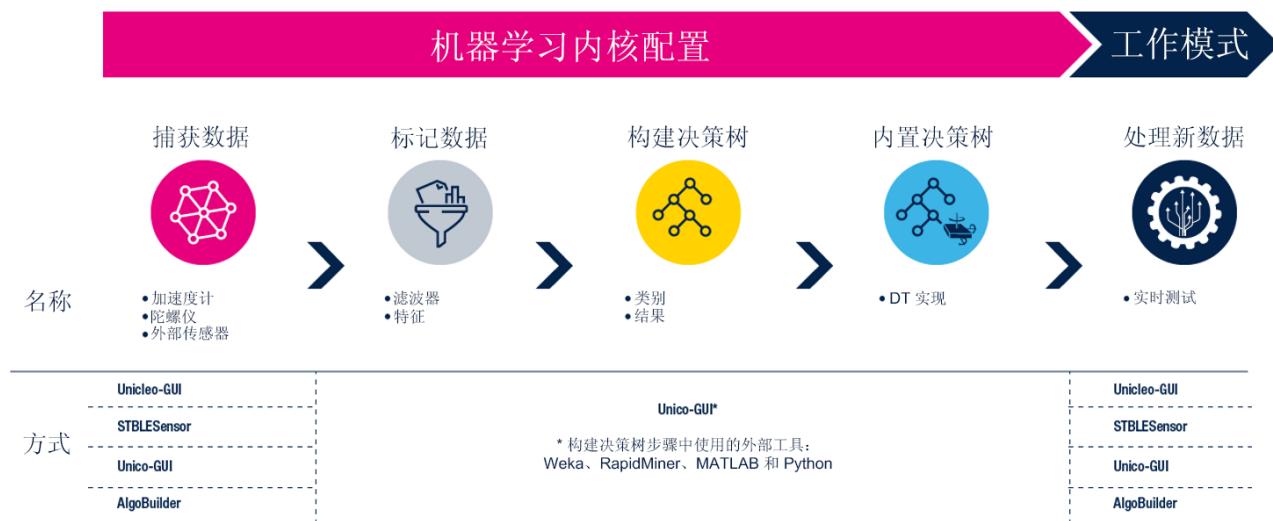
本文档提供有关 **LSM6DSV16X** 中可用的机器学习内核功能的信息。机器学习处理能力允许将一些算法从应用处理器转移到 **MEMS** 传感器，从而持续降低功耗。

通过决策树逻辑获得机器学习处理能力。决策树是由一系列可配置节点组成的数学工具。每个节点均采用“如果-则-否则”条件，并根据阈值评估输入信号（由传感器数据计算出的统计参数表示）。

可通过配置 **LSM6DSV16X** 来同时独立运行多达 4 个决策树。决策树存储在器件中，并在专用输出寄存器中生成结果。

可通过应用处理器随时读取决策树的结果。此外，可针对决策树结果的每次更改产生中断。

图 1. 机器学习内核监督式方法



1 LSM6DSV16X 中的机器学习内核

机器学习内核（以及有限状态机）是 LSM6DSV16X 中的主要嵌入式功能之一。它由能够在传感器本身中实现算法的一组可配置的参数和决策树组成。

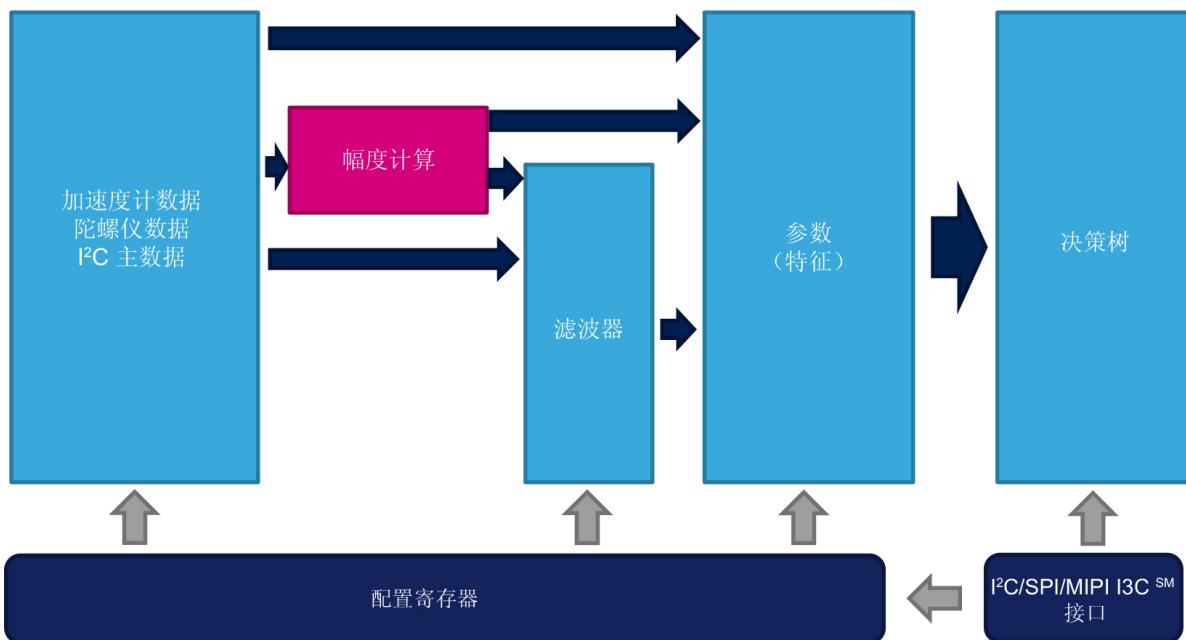
适用于机器学习内核的算法为可通过归纳法实现的算法，这涉及从观察结果搜索模式。采用该方法的一些算法示例包括：活动识别、健身活动识别、运动强度检测、振动强度检测、携带位置识别、情景感知、假阳性拒绝等。

机器学习内核的理念为使用加速度计、陀螺仪和外部传感器数据（可通过 I²C 主接口读取）来计算一组可由用户在定义的时间窗口中选择的统计参数（例如均值、方差、能量、峰值、过零信号等）。除了传感器输入数据以外，还可以通过应用器件中可用的一些可配置滤波器来定义一些滤波输入。

机器学习内核参数被称为“功能”，可用作可配置决策树的输入，该决策树可存储在器件中。

可存储在 LSM6DSV16X 中的决策树为由一系列节点组成的二叉树。在每个节点中，根据阈值评估统计参数（特征），以建立下一个节点中的演化。当达到叶时（树中的最后一个节点之一），决策树产生结果，可通过专用器件寄存器来读取该结果。

图 2. LSM6DSV16X 中的机器学习内核



可将机器学习内核输出数据配置为 15 Hz 到 240 Hz 之间的四个可用速率之一。嵌入式功能寄存器 EMB_FUNC_ODR_CFG_C (60h) 中的 MLC_ODR 位允许选择四个可用速率之一，如下表所示。

表 1. 机器学习内核输出数据速率

EMB_FUNC_ODR_CFG_C (60h) 中的 MLC_ODR 位	机器学习内核输出数据速率
000	15 Hz
001	30 Hz (默认值)
010	60 Hz
011	120 Hz
100	240 Hz

为实现 LSM6DSV16X 的机器学习处理能力，必须使用“监督式学习”方法，该方法包括：

- 确定一些要识别的类别；
- 为每个类别收集多个数据日志；
- 通过收集的日志执行一些数据分析，以学习通用规则，该规则允许将输入（数据日志）映射到输出（要识别的类别）。

例如，在活动识别算法中，要识别的类别可能为：静止、行走、慢跑、骑自行车、驾驶等。必须为每个类别获取多个数据日志，例如，多人执行相同的活动。

分析收集的数据日志的目的为：

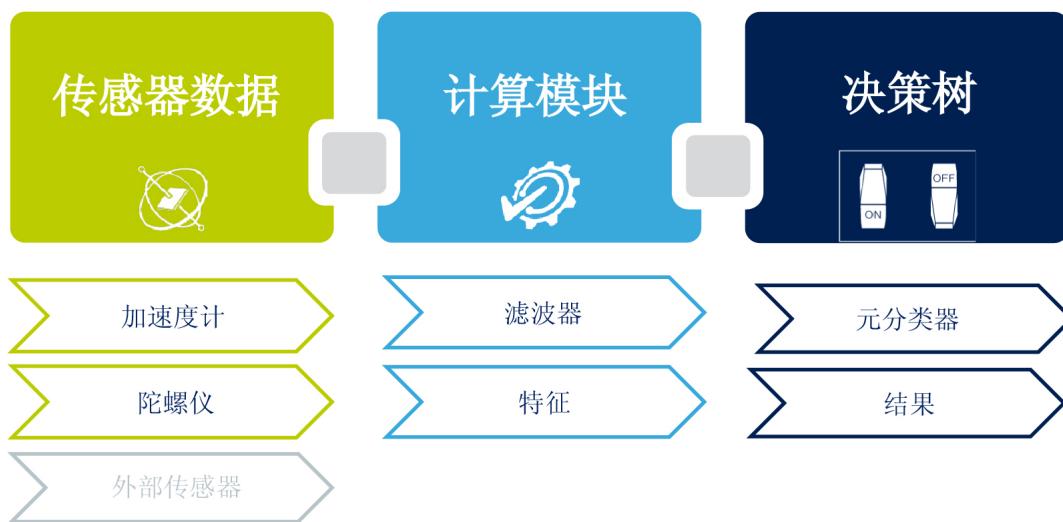
- 定义用于对不同的类别进行正确分类的功能；
- 定义将应用于输入数据的滤波器，以使用所选特征提高性能；
- 生成能够识别不同类别之一的专用决策树（将输入映射到输出）。

一旦定义了决策树，即可通过意法半导体提供的软件工具（如第 2 节 机器学习内核工具中所述）生成器件配置。决策树将在器件上运行，以最大限度地降低功耗。

要进一步详细了解 LSM6DSV16X 中的机器学习内核功能，可以将其视为三个主要模块（图 3）：

1. 传感器数据
2. 计算模块
3. 决策树

图 3. 机器学习内核模块



第一个模块被称为“传感器数据”，该数据包括内置于器件的加速度计和陀螺仪的数据，或可能通过 I^C 主接口（传感器集线器）连接到 LSM6DSV16X 的附加外部传感器的数据。

第一个模块中定义的机器学习内核输入用于第二个模块，即“计算模块”，可在其中应用滤波器和特征。特征为定义的时间窗口中的输入数据（或滤波后的数据）计算出的统计参数，用户可选择这些参数。

在计算模块中计算出的特征将用作机器学习内核的第三个模块的输入。该模块被称为“决策树”，这包括二叉树，该二叉树用于评估通过输入数据计算出的统计参数。在二叉树中，通过将这些参数与特定阈值进行比较来生成结果（在上述活动识别示例中，结果为：静止、行走、慢跑、骑自行车等）。也可以通过被称为“元分类器”的可选滤波器过滤决策树结果。机器学习内核结果将为决策树结果，其中包括可选的元分类器。

机器学习内核存储器以“动态”或“模块化”方式组织，以便最大限度地增加可在器件（滤波器、功能等）中配置的计算模块的数量。已设计用于生成 LSM6DSV16X 的配置的专用工具，以自动管理存储器使用。该工具在 Unico GUI 中可用，稍后将在第 2 节 机器学习内核工具中介绍。

以下几节详细说明了图 3 中所述的 LSM6DSV16X 中的机器学习内核的三个主要模块。

1.1 输入

LSM6DSV16X 用作组合（加速度计 + 陀螺仪）传感器，可生成加速度和角速率输出数据。加速度和角速率的 3 轴数据可用作机器学习内核的输入。图 4 和图 5 显示了加速度计和陀螺仪数字链中的机器学习内核模块输入。对于器件数据手册中所述的全部连接模式，机器学习内核 (MLC) 模块在两个数字链中的位置相同。

提示

作为 MLC 模块输入的数据在输出寄存器中具有更高的分辨率。

图 4. MLC 输入（加速度计）

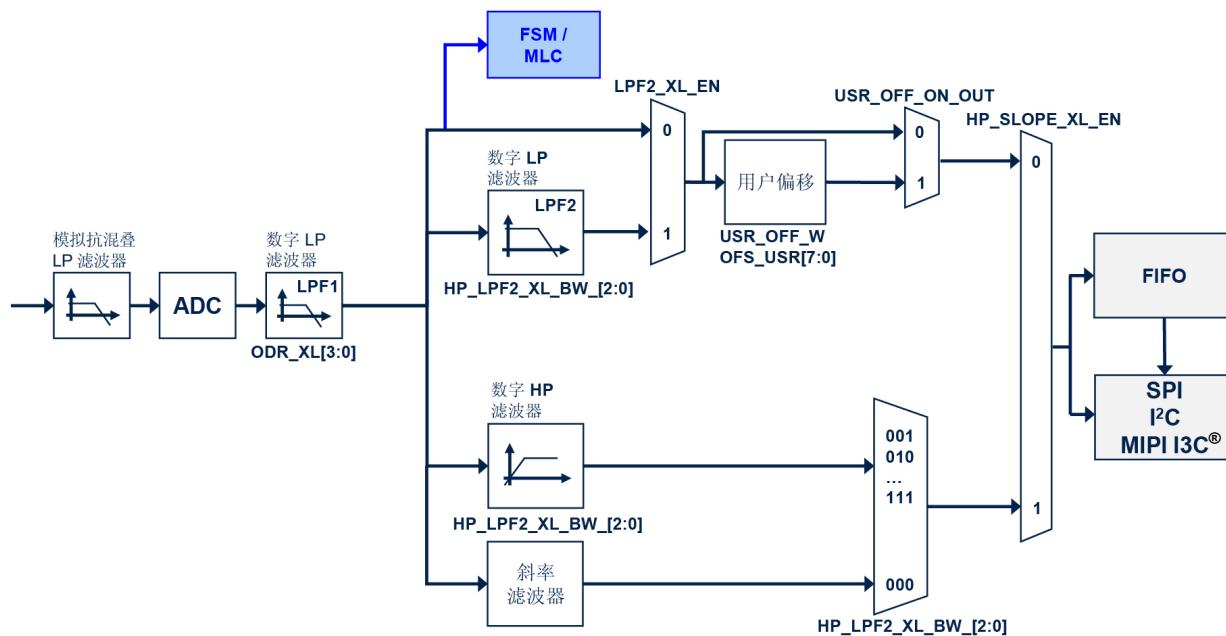
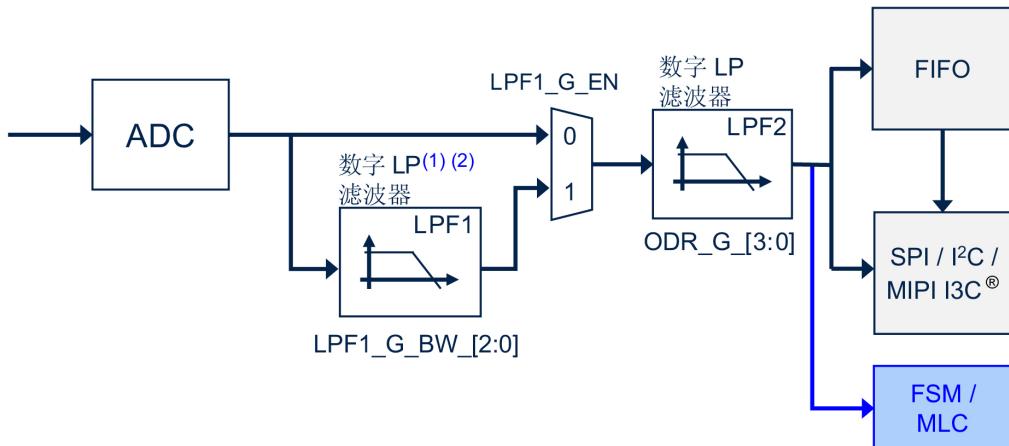


图 5. MLC 输入（陀螺仪）



输入数据的速率必须等于或高于嵌入式功能寄存器 EMB_FUNC_ODR_CFG_C (60h) 配置的机器学习内核数据速率，如表 1 中所述。

示例：在以 30 Hz 运行的活动识别算法中，必须选择 30 Hz 的机器学习内核 ODR，而传感器 ODR 必须等于或高于 30 Hz。

机器学习内核使用以下单位：

- 加速度计数据的单位为 [g]
- 陀螺仪数据的单位为 [rad/sec]
- 磁力计外部传感器数据的单位为 [gauss]

由于可通过传感器集线器功能（模式 2）将外部传感器（如磁力计）连接到 LSM6DSV16X，因此来自外部传感器的数据也可以用作机器学习处理的输入。在这种情况下，将前六个传感器集线器字节（每轴两个字节）视为 MLC 的输入。换句话说，MLC 直接将 LSM6DSV16X 寄存器 SENSOR_HUB_1 (02h) 到 SENSOR_HUB_6 (07h) 的内容作为输入。

使用外部传感器时，必须通过寄存器 MLC_EXT_SENSITIVITY_L (E8h) 和 MLC_EXT_SENSITIVITY_H (E9h) 设置外部传感器的灵敏度。

例如：对于诸如 LIS2MDL 的磁力计，灵敏度为 1.5 mG/LSB，将其转换为高斯时，灵敏度变为 0.0015 G/LSB，将其转换为 HFP（LSM6DSV16X 灵敏度寄存器的半精度浮点值）时，灵敏度变为 1624h。

灵敏度 [mG/LSB]	灵敏度 [G/LSB]	灵敏度 HFP
1.5 mG/LSB	0.0015 G/LSB	1624h

提示

半精度浮点格式表示为：

SEEEEEEFFFFFFF (*S*: 1 个符号位; *E*: 5 个指数位; *F*: 10 个分数位)。

以下程序允许更改外部磁力计数据的转换系数：

1. 将 80h 写入寄存器 01h // 使能对嵌入式功能寄存器的访问
2. 将 40h 写入寄存器 17h // PAGE_RW (17h) = 40h: 使能写事务
3. 将 11h 写入寄存器 02h // PAGE_SEL (02h) = 11h
4. 将 E8h 写入寄存器 08h // PAGE_ADDRESS (08h) = E8h
5. 将 [LSB] 转换系数（LIS2MDL 示例，24h）写入寄存器 09h
6. 将 11h 写入寄存器 02h // PAGE_SEL (02h) = 11h
7. 将 E9h 写入寄存器 08h // PAGE_ADDRESS (08h) = E9h
8. 将 [MSB] 转换系数（LIS2MDL 示例，16h）写入寄存器 09h
9. 将 00h 写入寄存器 17h // PAGE_RW (17h) = 00h: 禁止读/写事务
10. 将 00h 写入寄存器 01h // 禁止对嵌入式功能寄存器的访问

机器学习内核工具（如第 2 节 机器学习内核工具中所述）生成的配置中包含上述更改外部传感器灵敏度的程序示例，因此用户只需在 GUI 中设置灵敏度值，软件就会将值转换为寄存器设置。

MLC 中可用作外部传感器数据的另一个输入为 Qvar。MLC 中的 Qvar 输入可用作外部传感器 X 轴数据（两个字节）。考虑到 Qvar 数据以 LSB 表示，外部传感器的灵敏度可以设置为 3C00h（即，1 对应的半精度浮点值）。

LSM6DSV16X 中的 MLC 还可以使用具有三字节数据的外部传感器（如压力传感器）。在这种情况下，可以使用嵌入高级功能页 2 中提供的一些专用寄存器，以便正确设置灵敏度和偏移：

- EXT_3BYTE_SENSITIVITY_L (02h)
- EXT_3BYTE_SENSITIVITY_H (03h)
- EXT_3BYTE_OFFSET_XL (06h)
- EXT_3BYTE_OFFSET_L (07h)
- EXT_3BYTE_OFFSET_H (08h)

EXT_3BYTE_SENSITIVITY 寄存器中的灵敏度值以 HFP（半精度浮点值）表示。EXT_3BYTE_OFFSET 寄存器中的偏移值以 LSB 表示。

汇总机器学习内核输入：

- 加速度计数据转换系数由器件自动处理。
- 陀螺仪数据转换系数由器件自动处理。
- 器件不自动处理外部传感器数据转换系数。用户必须设置转换系数，以使机器学习内核能够使用正确的测量单位。

适用于所有传感器数据（加速度计、陀螺仪和外部传感器）的附加输入都是范数。机器学习内核（位于 LSM6DSV16X 中）通过 3 轴数据在内部计算范数和平方范数。这两个附加信号可用作机器学习处理的输入。

使用以下公式计算输入数据的范数和平方范数：

$$V = \sqrt{x^2 + y^2 + z^2}$$
$$V^2 = x^2 + y^2 + z^2$$

范数与平方范数数据可用于决策树，以确保为用户提供高水平的程序定制。

提示 MLC 输入的数据速率通过 *MLC_ODR* 位设置。如果传感器 ODR 高于 *MLC_ODR*，MLC 将自动抽取输入数据（无需任何额外的滤波）。

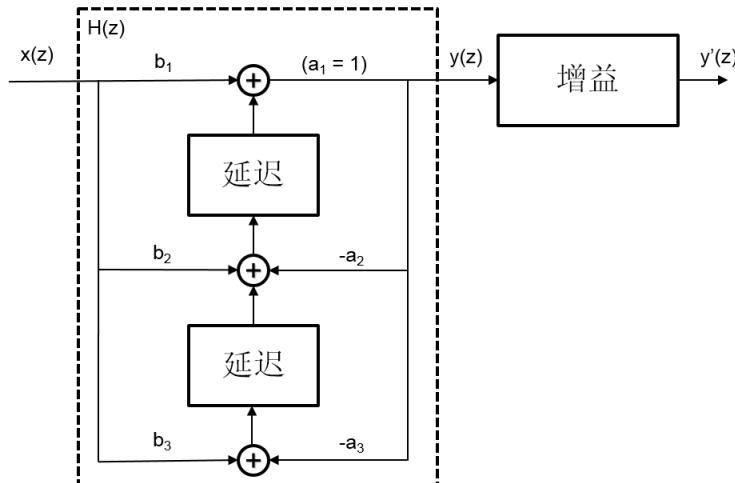
建议选择等于传感器 ODR 的 *MLC_ODR*，以避免抽取 MLC 输入。此外，也可以选择低于传感器 ODR 的 *MLC_ODR*，但频率响应会有所不同，可能导致 MLC 解决方案的精度降低。

1.2 滤波器

可通过机器学习内核逻辑中可用的各种滤波器来过滤上一节中的输入数据。机器学习内核滤波的基本元素为二阶 IIR 滤波器，如图 6 所示。

提示 MLC 模块中可用的滤波器与器件上可用的任何其他滤波器无关（本节所述的滤波器将在图 4 和图 5 的 MLC 模块中进行说明）。

图 6. 滤波器基本元素



通用 IIR 2 阶滤波器的传递函数如下：

$$H(z) = \frac{b_1 + b_2 z^{-1} + b_3 z^{-2}}{1 + a_2 z^{-1} + a_3 z^{-2}}$$

根据图 6，可将输出定义为：

$$\begin{aligned} y(z) &= H(z) \cdot x(z) \\ y'(z) &= y(z) \cdot Gain \end{aligned}$$

为优化内存使用，机器学习内核具有针对不同类型的滤波器（高通、带通、IIR1 和 IIR2）的默认系数。在选择滤波器类型后，机器学习内核工具将通过询问所需的滤波器系数来帮助配置滤波器。下表显示了取决于所选滤波器类型的系数默认值和可配置值。通过设置不同的系数，可针对特定应用调整滤波器。

表 2. 滤波系数

滤波器类型/系数	b ₁	b ₂	b ₃	a ₂	a ₃	增益
高通滤波器	0.5	-0.5	0	0	0	1
带通滤波器	1	0	-1	可配置	可配置	可配置
IIR1 滤波器	可配置	可配置	0	可配置	0	1
IIR2 滤波器	可配置	可配置	可配置	可配置	可配置	1

滤波系数值以半精度浮点格式表示：SEEEEEFFFFFFFFF (S: 1 个符号位；E: 5 个指数位；F: 10 个分位数)。

1.2.1 滤波系数

可使用不同的工具（包括 MATLAB®、Octave 和 Python）计算 IIR 滤波器系数。例如，在 MATLAB® 中，以下函数可用于生成低通滤波器的系数：

```
[b, a] = butter( N, f_cut / (ODR/2), 'low' )
```

其中：

- N 为 IIR 滤波器的阶数（IIR1 为 1, IIR2 为 2）
- f_{cut} 为滤波器的截止频率 [Hz]
- ODR 为机器学习内核数据速率 [Hz]
- “low”（或“high”）为要实施的滤波器类型（低通或高通）

提示

可配置一个高通滤波器，其截止频率为带宽 (ODR/4) 的一半，而无需插入系数。机器学习内核具有一些针对该配置的预定义系数。

以下函数允许通过 MATLAB® 生成带通滤波器系数：

```
[b, a] = butter(1, [f1 f2] / (ODR/2), 'bandpass')
```

提示

由于带通滤波器仅可配置 a_2 、 a_3 和增益，因此应通过设置增益 = $b(1)$ 来标准化 b 向量。带通滤波器在 MATLAB® 和 Python 中以一阶滤波器的形式生成。

示例：

$b = [0.2929 \ 0 \ -0.2929]; \ a = [1.0 \ -0.5858 \ 0.4142];$

可写成 $b = [1 \ 0 \ -1]$ ，增益 = 0.2929。

因此，带通滤波器的系数为：

$a_2 = -0.5858; \ a_3 = 0.4142;$ 增益 = 0.2929。

表 3 显示了一些滤波器系数的示例（大多数示例考虑了 30 Hz 的 ODR）。

设计高通和带通 IIR 滤波器时，应考虑采用半精度浮点值的滤波器的稳定性。如果滤波器不是很稳定，则分辨率会有所降低，进而导致信号出现一定偏差。

建议在截止频率（标准化）低于 0.02 [2*f_cutoff/ODR] 时使用一阶 IIR 滤波器，或者增大截止频率。

表 3. 滤波器系数示例

滤波器类型/系数	b_1	b_2	b_3	a_2	a_3	增益
高通 IIR1, $f_{cut} = 1$ Hz, ODR = 30 Hz	0.905	-0.905	-	-0.8096	-	1
高通 IIR1, $f_{cut} = 2$ Hz, ODR = 30 Hz	0.8247	-0.80247	-	-0.6494	-	1
高通 IIR1, $f_{cut} = 5$ Hz, ODR = 30 Hz	0.634	-0.634	-	-0.268	-	1
高通 IIR1, $f_{cut} = 10$ Hz, ODR = 30 Hz	0.366	-0.366	-	0.268	-	1
高通 IIR2, $f_{cut} = 1$ Hz, ODR = 30 Hz	0.8623	-1.725	0.8623	-1.705	0.7437	1
高通 IIR2, $f_{cut} = 2$ Hz, ODR = 30 Hz	0.743	-1.486	0.743	-1.419	0.553	1
高通 IIR2, $f_{cut} = 5$ Hz, ODR = 30 Hz	0.465	-0.93	0.465	-0.62	0.2404	1
高通 IIR2, $f_{cut} = 10$ Hz, ODR = 30 Hz	0.155	-0.31	0.155	0.62	0.2404	1
低通 IIR1, $f_{cut} = 1$ Hz, ODR = 30 Hz	0.0951	0.0951	-	-0.8096	-	1

滤波器类型/系数	b ₁	b ₂	b ₃	a ₂	a ₃	增益
低通 IIR1, f _{cut} = 2 Hz, ODR = 30 Hz	0.1753	0.1753	-	-0.6494	-	1
低通 IIR1, f _{cut} = 5 Hz, ODR = 30 Hz	0.366	0.366	-	-0.268	-	1
低通 IIR1, f _{cut} = 10 Hz, ODR = 30 Hz	0.634	0.634	-	0.268	-	1
低通 IIR2, f _{cut} = 1 Hz, ODR = 30 Hz	0.00953	0.01906	0.00953	-1.705	0.7437	1
低通 IIR2, f _{cut} = 2 Hz, ODR = 30 Hz	0.03357	0.06714	0.03357	-1.419	0.553	1
低通 IIR2, f _{cut} = 5 Hz, ODR = 30 Hz	0.155	0.31	0.155	-0.62	0.2404	1
低通 IIR2, f _{cut} = 10 Hz, ODR = 30 Hz	0.465	0.93	0.465	0.62	0.2404	1
带通 IIR2, f ₁ = 1.5 Hz, f ₂ = 5 Hz, ODR = 30 Hz	1	0	-1	-1.203	0.4453	0.2773
带通 IIR2, f ₁ = 0.2 Hz, f ₂ = 1 Hz, ODR = 120 Hz	1	0	-1	-1.958	0.959	0.02052

1.3 特征

这些特征为通过机器学习内核输入计算出的统计参数。可用于特征计算的机器学习内核输入如下：

- 传感器输入数据，包括
 - X、Y、Z 轴的传感器数据（如 Acc_X、Acc_Y、Acc_Z、Gyro_X、Gyro_Y、Gyro_Z）
 - 外部传感器数据（如 ExtSens_X、ExtSens_Y、ExtSens_Z）
 - 传感器/外部传感器数据的范数和平方范数信号（Acc_V、Acc_V2、Gyro_V、Gyro_V2、ExtSens_V、Ext_Sens_V2）
- 滤波数据（如 Acc_Z 上的高通、Acc_V2 上的带通等）

所有特征均在定义的时间窗口内计算，这也被称为“窗口长度”，因为它以样本数表示。窗口大小必须由用户确定，这对于机器学习处理非常重要，因为决策树中的所有统计参数均在该时间窗口中评估。这不是移动窗口，对于每个 WL 样本，只计算一次特征（其中，WL 为窗口大小）。

窗口长度值可能为 1 到 255 个样本。窗口长度值的选择取决于 MLC 数据速率（嵌入式功能寄存器 EMB_FUNC_ODR_CFG_C (60h) 中的 MLC_ODR 位），这将引入用于产生机器学习内核结果的等待时间，同时也取决于特定应用程序或算法。例如，在活动识别算法中，可决定每隔 2 秒或 3 秒计算一次特征，这意味着，考虑到传感器以 30 Hz 运行，窗口长度应分别约为 60 或 90 个样本。

评估机器学习内核中的某些特征需要一些附加参数（如附加阈值）。下表显示了机器学习内核中的所有可用特征，包括附加参数。

提示

可以在 MLC 中配置的最大特征数为 31。特征值限制在 ±65536 范围内。

表 4. 特征

特征	附加参数
均值	-
方差	-
能量	-
峰峰值	-
过零信号	阈值
正过零信号	阈值
负过零信号	阈值
峰值探测器	阈值
正峰值探测器	阈值
负峰值探测器	阈值
最小值	-
最大值	-
递归均值	b ₁ 、b ₂ 、a ₂
递归 RMS	b ₁ 、b ₂ 、a ₂
递归方差	b ₁ 、b ₂ 、a ₂
递归最大值	ths、c_start
递归最小值	ths、c_start
递归峰峰值	ths、c_start

1.3.1 均值

“均值” 特征通过以下公式计算定义的时间窗口 (WL) 中的所选输入 (I) 的均值：

$$\text{均值} = \frac{1}{WL} \sum_{k=0}^{WL-1} I_k$$

1.3.2 方差

“方差” 特征通过以下公式计算定义的时间窗口 (WL) 中的所选输入 (I) 的方差：

$$\text{方差} = \left(\frac{\sum_{k=0}^{WL-1} I_k^2}{WL} \right) - \left(\frac{\sum_{k=0}^{WL-1} I_k}{WL} \right)^2$$

1.3.3 能量

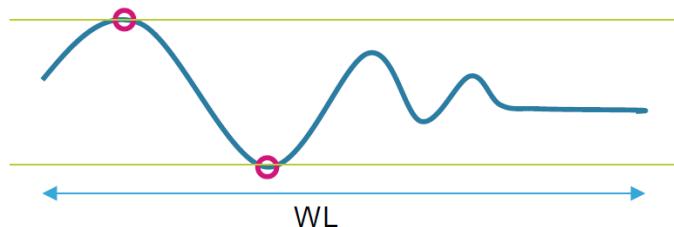
“能量” 特征通过以下公式计算定义的时间窗口 (WL) 中的所选输入 (I) 的能量：

$$\text{能量} = \sum_{k=0}^{WL-1} I_k^2$$

1.3.4 峰峰值

“峰峰值” 特征可计算定义的时间窗口中的所选输入的最大峰峰值。

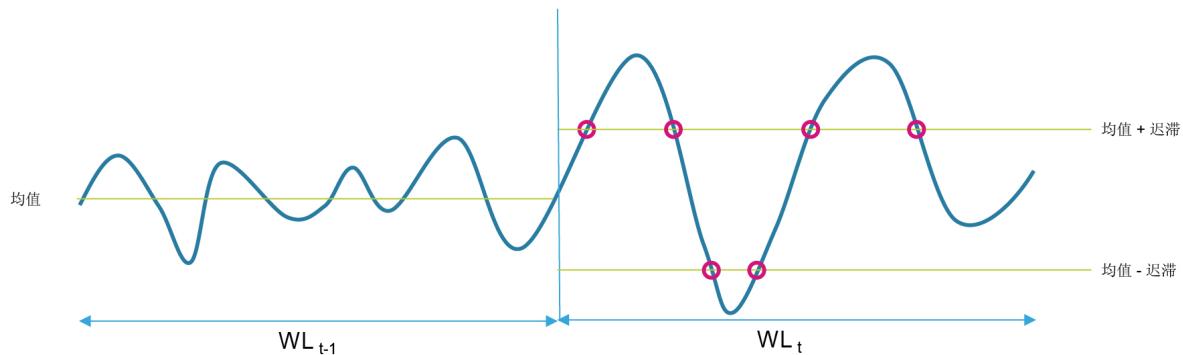
图 7. 峰峰值



1.3.5 过零信号

“过零信号”特征可计算所选输入跨过某个阈值的次数。该内部阈值定义为在前一个窗口中计算的均值（“均值”特征）与用户定义的迟滞之和。

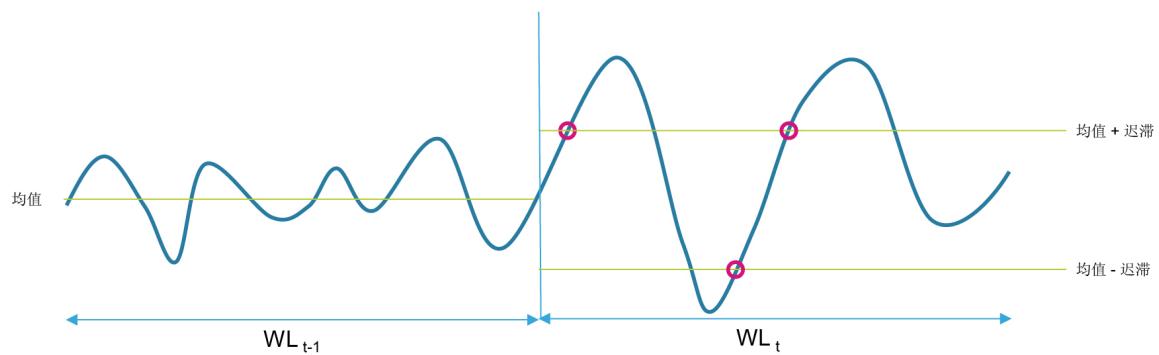
图 8. 过零信号



1.3.6 正过零信号

“正过零信号”特征可计算所选输入跨过某个阈值的次数。该内部阈值定义为在前一个窗口中计算的均值（“均值”特征）与用户定义的迟滞之和。此特征仅考虑具有正斜率的过渡。

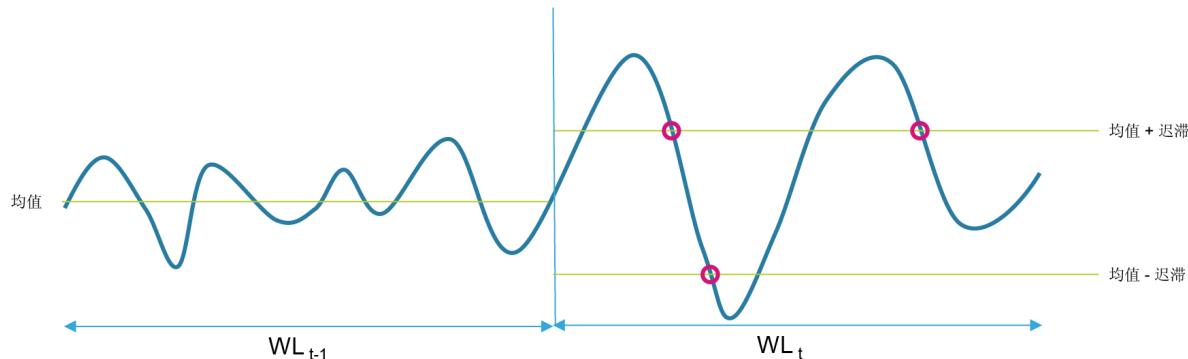
图 9. 正过零信号



1.3.7 负过零信号

“负过零信号”特征可计算所选输入跨过某个阈值的次数。该内部阈值定义为在前一个窗口中计算的均值（“均值”特征）与用户定义的迟滞之和。此特征仅考虑具有负斜率的过渡。

图 10. 负过零信号



1.3.8 峰值探测器

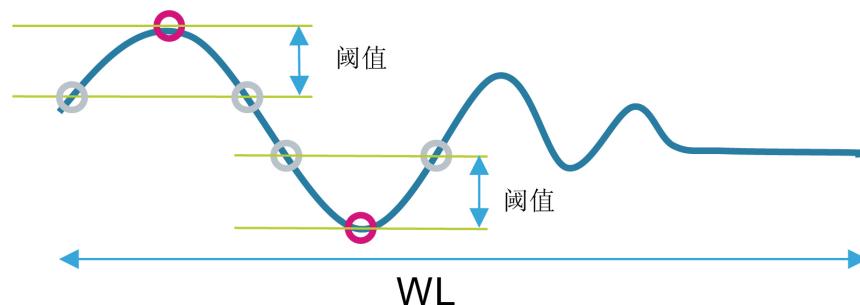
“峰值探测器”特征可计算定义的时间窗口中的所选输入的峰值数（正负）。

用户必须为此特征定义一个阈值，并考虑使用三个值的缓冲区进行评估。如果三值缓冲区的第二个值大于（或小于）所选阈值的其他两个值，则峰值数会增加。

为计算此特征而考虑的三个值的缓冲区是时间窗口内的移动缓冲区。

下图显示了此特征的计算示例，已在时间窗口中检测到两个峰值（一正一负）。

图 11. 峰值探测器



1.3.9 正峰值探测器

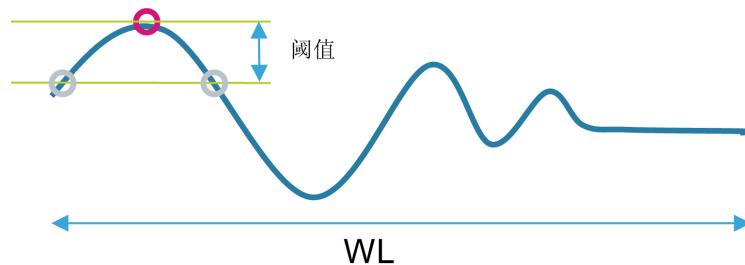
“正峰值探测器”特征可计算定义的时间窗口中的所选输入的正峰值数量。

用户必须为此特征定义一个阈值，并考虑使用三个值的缓冲区进行评估。如果三值缓冲区的第二个值大于所选阈值的其他两个值，则峰值数会增加。

为计算此特征而考虑的三个值的缓冲区是时间窗口内的移动缓冲区。

下图显示了此特征的计算示例，仅在时间窗口中检测到一个峰值（正）。

图 12. 正峰值探测器



1.3.10 负峰值探测器

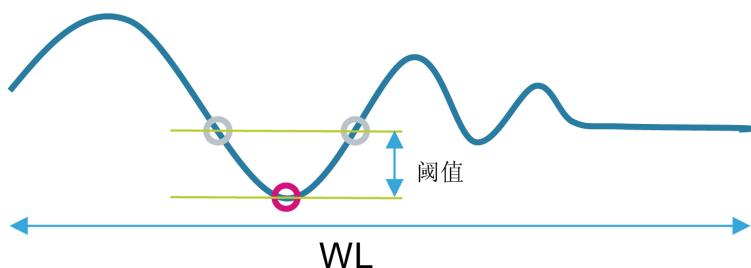
“负峰值探测器”特征可计算定义的时间窗口中的所选输入的负峰值数量。

用户必须为此特征定义一个阈值，并考虑使用三个值的缓冲区进行评估。如果三值缓冲区的第二个值小于所选阈值的其他两个值，则峰值数会增加。

为计算此特征而考虑的三个值的缓冲区是时间窗口内的移动缓冲区。

下图显示了此特征的计算示例，仅在时间窗口中检测到一个峰值（负）。

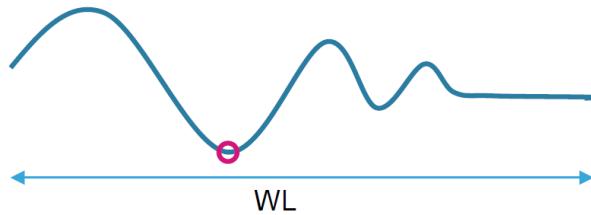
图 13. 负峰值探测器



1.3.11 最小值

“最小值” 特征可计算定义的时间窗口中的所选输入的最小值。下图显示了时间窗口中的最小值示例。

图 14. 最小值

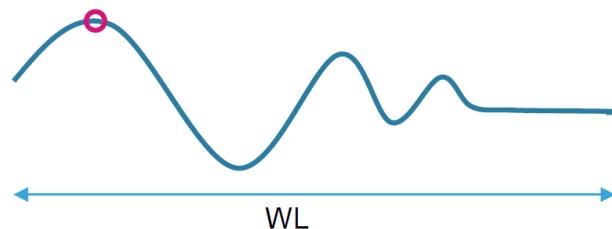


1.3.12 最大值

“最大值” 特征可计算定义的时间窗口中的所选输入的最大值。

下图显示了时间窗口中的最大值示例。

图 15. 最大值



1.3.13 递归特征

除了上述标准特征之外，**LSM6DSV16X** 还有一组称为“递归”的高级特征。

递归算法的计算复杂度低，存储需求极少，因此广泛用于实时嵌入式系统。

在 **MLC** 中，每个窗口长度（非重叠窗口）计算的特征都会存在一些限制，但递归特征可以克服这些限制，从而更好地检测较短事件，简化不同算法的共存问题。

与窗口化特征不同，递归特征：

- 在每个 **MLC_ODR** 提供一个有效值
- 具有无限窗口（无限冲激响应）
- 可针对使用权重的特征调整“存储时间”

MLC 中的递归特征分为以下两组：

- 一阶均值、**RMS**、方差
- 最大值、最小值、峰峰值

1.3.13.1 递归均值、**RMS**、方差

均值、**RMS** 和方差的递归特征基于 **IIR** 低通滤波器。这些特征的参数为 b_1 、 b_2 和 a_2 。示例： $b_1 = 1$ ； $b_2 = 0.25$ ； $a_2 = 0.75$

下表列出了第一组递归特征的三个公式。

表 5. 第一组递归特征

递归特征	公式
递归均值	$IIR_{LP}[x_i]$
递归方差	$IIR_{LP}[x_i^2] - IIR_{LP}^2[x_i]$
递归 RMS	$\sqrt{IIR_{LP}[x_i^2]}$

1.3.13.2 递归最大值、最小值和峰峰值

最大值、最小值和峰峰值的递归特征基于与包络探测器相似的概念：

- 最大值/最小值高于/低于当前值时，将等于输入值
- 否则，最大值/最小值将以可配置系数开始衰减

下表列出了第二组递归特征的三个公式。

这些特征的可配置参数包括：

- 阈值，用于设置“复位值”（应设置为输入信号的平均值）
- c_{start} ，衰减系数的初始值，必须设置为 0 和 1 之间的值（值越小，更新速度越快）

表 6. 第二组递归特征

递归特征	公式	
递归最大值 (Max_i) ⁽¹⁾	$Max_i = in_i$	如果 $in_i > Max_{i-1}$ $C_{max} = C_{start}$
	$Max_i = THS + (Max_{i-1} - THS) \cdot c_{max}$	如果 $in_i \leq Max_{i-1}$ $C_{max} = C_{start} \cdot c_{max}$
递归最小值 (Min_i) ⁽²⁾	$Min_i = in_i$	如果 $in_i < Min_{i-1}$ $C_{min} = C_{start}$
	$Min_i = THS - (THS - Min_{i-1}) \cdot c_{min}$	如果 $in_i \geq Min_{i-1}$ $C_{min} = C_{start} \cdot c_{min}$
递归峰峰值	$pk2pk_i = Max_i - Min_i$	

1. 递归最大值 (Max_i) 不得低于阈值 (THS)。

2. 递归最小值 (Min_i) 不得高于阈值 (THS)。

1.3.14 特征选择

根据特定应用选择将用于机器学习内核配置的特征。

考虑到使用过多特征可能导致决策树过拟合和过大，建议先通过选择四个最常用的特征来启动：

- 均值
- 方差
- 能量
- 峰峰值

如果这些特征的性能不佳，并且为了提高精度，可以考虑使用其他特征来更好地分离类别。

此外，还必须根据特定应用选择特征计算的输入数据（来自加速度计、陀螺仪）和轴（如 X、Y、Z、V）。某些类别与传感器方向密切相关（即，使用设备携带位置的应用），因此最好使用独立轴（X, Y, Z）。其他类别（例如行走）与方向无关，因此最好使用范数（V 或 V2）。

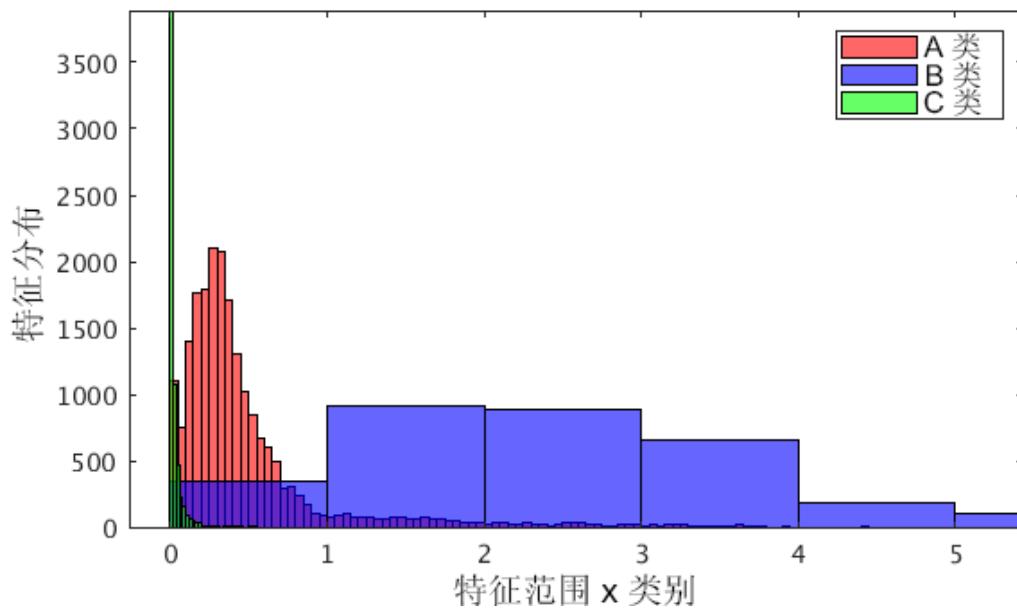
有时，基本特征（均值、方差、能量等）可能无法帮助区分主导频率，因此可以使能嵌入式数字滤波器来选择特定频率区域。使用滤波信号可以更精确地区分某些类别。例如，如果用户正在行走，则典型信号约为 1-2 Hz，而如果用户正在慢跑，则典型信号约为 2.5-4 Hz。

通过测试不同类别之间的分离程度，可以评估单一特征的信息贡献度。此分析可采用图形方式来完成，具体方法为按以下示例所述绘制 1D/2D 图。

1.3.14.1 单个特征的直方图 (1D 图)

下图显示了单个特征的三个不同类别计算值的直方图。这三个类别的分布合理分散，因此预计可通过这一特征获得重要的信息。该特征的类别计算精度约为 75%（供参考）。

图 16. 单个特征的三个不同类别的分布



1.3.14.2 显示两个特征 (2D 图)

下图显示了与双类别分类问题相关的 2D 图，其中选择了两个特征：

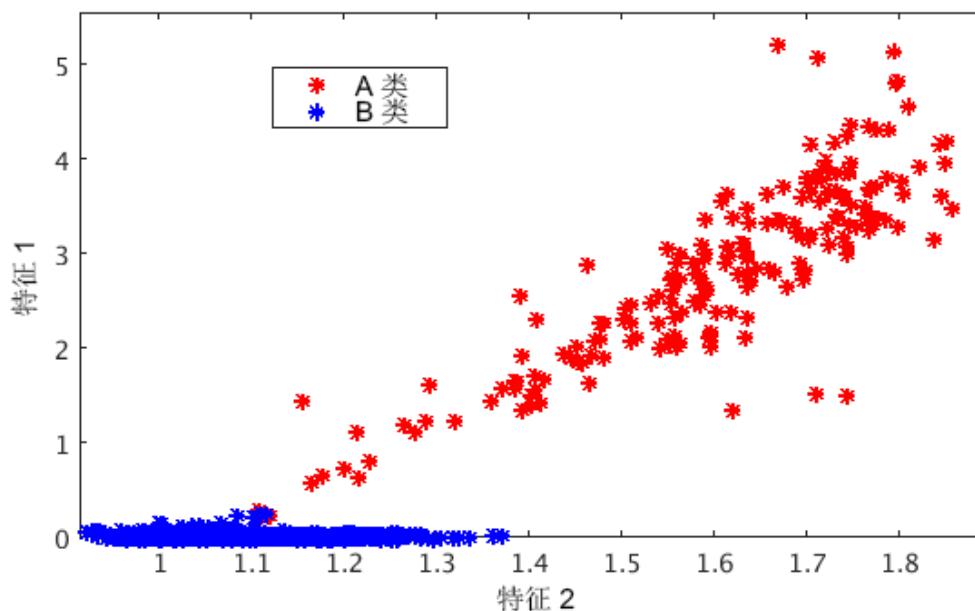
- 图中的纵轴为特征 1
- 图中的横轴为特征 2

在本例中，两个类别明显严格分离：

- A 类用红色表示
- B 类用蓝色表示

将两个特征相结合可获得理想的信息贡献度。本例得到的类别精度大于 95%（供参考）。

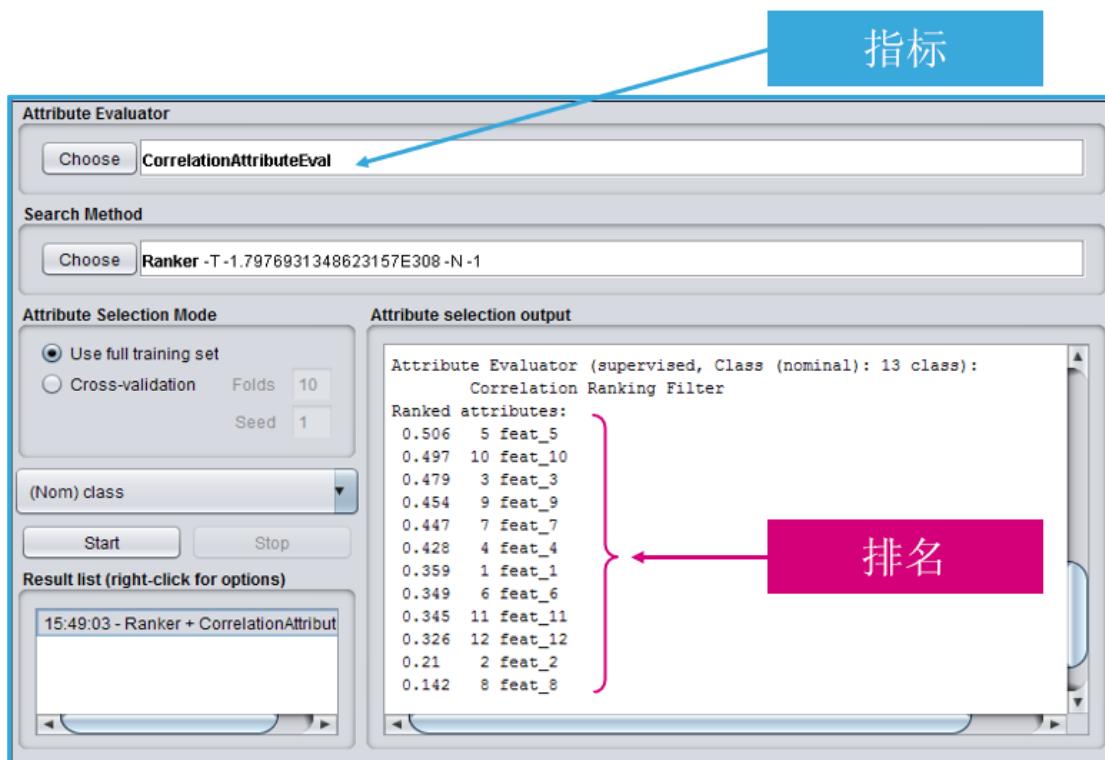
图 17. 显示两个特征和两个类别



1.3.14.3 特征排名

可以使用各种机器学习工具，根据各个特征的信息贡献度自动对特征进行排名。这种形式的输出排名基于相关性、信息增益、概率距离、熵等标准/指标。下面以 Weka 工具为例，该工具会自动处理生成最优决策树所需的计算，具体如下图所示。

图 18. 使用自动输出工具实现的排名



请注意，不同特征可能共享同一信息贡献度。对此，可以进行二次评估，具体方法为显示单个特征或者逐个检查各个特征子集可获得的精度以及所有特征组合在一起可获得的精度，如前几节所述。

最终可以考虑一下所选的特征数。通常，所选的特征数越多：

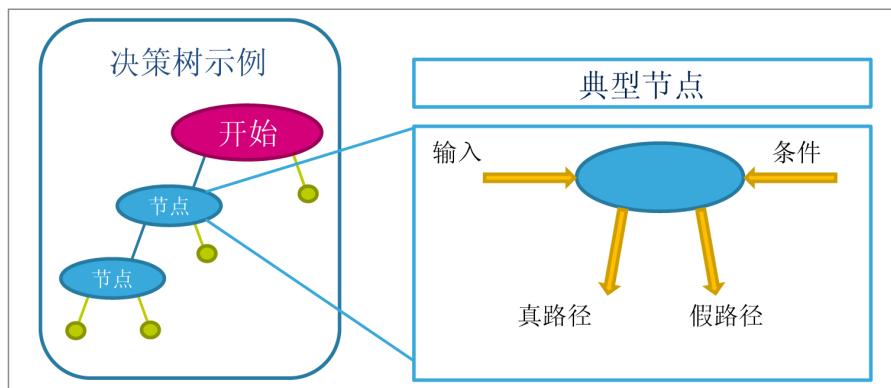
- 过拟合的风险越高
- 生成的决策树越大

1.4 决策树

决策树是根据训练数据构建的预测模型，训练数据可以存储在 LSM6DSV16X 中。训练数据是为要识别的每个类别获取的数据日志（在活动识别示例中，类别可能为行走、慢跑、驾驶等）。

前面各节中所述的计算模块为决策树的输入。决策树的每个节点都包含一个条件，将在该条件下通过特定阈值来评估特征。如果条件为真，则评估真路径中的下一个节点。如果条件为假，则评估假路径中的下一个节点。决策树的状态将逐节点演化，直至找到结果为止。决策树的结果是在数据收集开始时定义的类别之一。

图 19. 决策树节点



决策树在每个时间窗口（用户为计算特征而设置的“窗口长度”参数）产生一个新结果。窗口长度以样本数来表示。将表示窗口长度的样本数除以选定的 MLC 数据速率 (MLC_ODR) 可以得到时间窗口：

$$\text{时间窗口} = \text{窗口长度} / \text{MLC_ODR}$$

例如，当窗口长度为 120 个样本、MLC 数据速率为 120 Hz 时，对应的时间窗口如下：

$$\text{时间窗口} = 120 \text{ 个样本} / 120 \text{ Hz} = 1 \text{ 秒}$$

也可以通过被称为“元分类器”的附加（可选）滤波器过滤决策树结果，如第 1.5 节 元分类器所述。

可通过 LSM6DSV16X 寄存器的嵌入式高级功能页 1 中的专用寄存器访问机器学习内核结果（过滤或未过滤决策树结果）（如表 7 中所示）。可连续读取（轮询）这些寄存器，以检查决策树输出。MLC_STATUS_MAINPAGE (38h) 寄存器包含 8 个可能的决策树的中断状态位。当相应的决策树值更改时，自动将这些位设为 1。此外，也可以通过设置 MLC_INT1 (0Dh) 寄存器将使用这些位生成的中断状态信号驱动到 INT1 引脚，或通过设置 MLC_INT2(11h) 寄存器驱动到 INT2 引脚（表 8）。当机器学习内核结果改变时，使用中断信号可唤醒执行其他任务或休眠（以节省功耗）的 MCU。

默认情况下，机器学习内核中断信号为脉冲信号。脉冲中断时长由机器学习内核、有限状态机和传感器 ODR 中的最快 ODR 定义：

$$\text{中断脉冲时长} = 1 / \max(\text{MLC_ODR}, \text{FSM_ODR}, \text{XL_ODR}, \text{GYRO_ODR})$$

也可以通过嵌入式功能寄存器 PAGE_RW (17h) 中的 EMB_FUNC_LIR 位将机器学习内核中断信号设为锁存。

表 7. 决策树结果

寄存器	内容
MLC1_SRC (70h)	决策树 1 的结果
MLC2_SRC (71h)	决策树 2 的结果
MLC3_SRC (72h)	决策树 3 的结果
MLC4_SRC (73h)	决策树 4 的结果

表 8. 决策树中断

寄存器	内容
MLC_STATUS_MAINPAGE (4Bh)	将更改的中断状态位包含在决策树结果中
MLC_STATUS (15h)	将更改的中断状态位包含在决策树结果中
MLC_INT1 (0Dh)	允许将决策树的中断状态位发送到 INT1 引脚 ⁽¹⁾
MLC_INT2 (11h)	允许将决策树的中断状态位发送到 INT2 引脚 ⁽²⁾

1. 如果 MD1_CFG (5Eh) 的 INT1_EMB_FUNC 位设置为 1，则使能发送。

2. 如果 MD2_CFG (5Fh) 的 INT2_EMB_FUNC 位设置为 1，则使能发送。

1.4.1 LSM6DSV16X 中的决策树限制

就决策树数量、决策树大小、决策树结果的数量而言，LSM6DSV16X 的机器学习内核资源有限。

最多可将 4 种不同的决策树存储在 LSM6DSV16X 中，但所有决策树的节点总数之和不得超过 128(*)。LSM6DSV16X 中的每个决策树最多可包含 16 个结果。

(*) 该数量可能受所配置的特征和滤波器数量限制。通常，如果使用较少的滤波器和特征，则对决策树的大小没有进一步限制。但是，如果使用许多滤波器和特征，则对决策树的最大节点数略有限制。该工具用于向用户通知决策树的可用节点。

下表总结了 LSM6DSV16X 的限制。

表 9. LSM6DSV16X 中的决策树限制

	LSM6DSV16X
最大决策树数量	4
最大节点数（所有决策树的总数）	128 (*)
每个决策树的最大结果数	16

提示

使用多个决策树时，前几节中描述的所有参数（输入、滤波器、在时间窗口中计算的特征、时间窗口本身以及数据速率）对于所有决策树都是通用的。

1.5 元分类器

元分类器是决策树输出滤波器。元分类器使用一些内部计数器来过滤决策树输出。

决策树输出可分为多个子组（例如，可在同一子组中管理相似的类别）。内部计数器可用于决策树输出的所有子组。当决策树的结果是子组类别之一时，特定子组的计数器增加，否则减小。当计数器达到定义的值（被称为“结束计数器”）（由用户设置）时，机器学习内核的输出将更新。结束计数器允许的值为 0 到 14。

表 10. 元分类器示例

决策树结果	A	A	A	B	A	B	B	B	A	B	B	B	B	A	A	A
计数器 A (结束计数器 = 3)	1	2	3	2	3	2	1	0	1	0	0	0	0	1	2	3
计数器 B (结束计数器 = 4)	0	0	0	1	0	1	2	3	2	3	4	4	4	3	2	
机器学习内核结果 (包括元分类器)	x	x	A	B	B	B	B	A								

上表显示了通过元分类器过滤决策树输出的效果。该表的第一行包含元分类器之前的决策树输出。计数器 A 和计数器 B 是两个决策树结果（“A”和“B”）的内部计数器。在活动识别示例中，结果“A”可能为行走，结果“B”可能为慢跑。当内部计数器“A”达到值 3（这是计数器“A”的结束计数器）时，将转换为结果“A”。当内部计数器“B”达到值 4 时，将转换为结果“B”。

元分类器的目的为减少假阳性，以避免生成仍不稳定的输出，并减少决策树结果的转换。

1.5.1 LSM6DSV16X 中的元分类器限制

元分类器具有数量有限的子组，可在 LSM6DSV16X 中使用 4 个子组。可能需要使用元分类器对同一子组中的相似类别进行分组。

表 11. LSM6DSV16X 中的元分类器限制

	LSM6DSV16X
每个决策树的最大结果数	16
每个决策树的元分类器结果子组	4

提示

可配置多个元分类器。一个元分类器可用于机器学习内核中配置的任何决策树。

1.6 有限状态机接口

LSM6DSV16X 还提供了一种可配置的有限状态机，该状态机适用于演绎算法，尤其是手势识别。

有限状态机和决策树可组合在一起，以提高运动检测精度。

可通过 LSM6DSV16X 中提供的有限状态机检查机器学习内核所产生的决策树结果；这可以通过 CHKDT 条件来实现（如应用笔记 AN5882 LSM6DSV16X：有限状态机所述）。有限状态机也可以使用 MLC 滤波器和特征，滤波器和特征的标识符在 Unico-GUI 生成的配置文件 (.ucf) 中提供（参见图 20）。

默认情况下，FSM 程序在 MLC 算法之前执行。不过，可以通过寄存器 EMB_FUNC_EN_A (04h) 中的 MLC_before_FSM 位调换顺序。

提示

MLC 和 FSM 共享某一部分存储空间。同时使用 MLC 和 FSM 时，务必使用 Unico-GUI 中集成的合并工具适当组合二者的配置。可以在 Unico 主窗口的 [Options] (选项) 选项卡中单击 [Advanced MLC+FSM...] (高级 MLC+FSM...) 按钮找到该工具。

1.7 FIFO 中的 MLC 数据

LSM6DSV16X 可将 MLC 数据存储到 FIFO 中：

- MLC 结果**: 通过 EMB_FUNC_FIFO_EN_A (44h) 中的 MLC_FIFO_EN 位, 可以使能将机器学习内核结果批量存储到 FIFO 缓冲区中 (每个 MLCx_SRC 寄存器存储一个 FIFO 条目)。在这种情况下, 使用标记 1Ah。
- MLC 滤波器**可以在 MLC (MLC_ODR) 处理每个样本时存储到 FIFO 中。每个轴会对应有一个数据字存储到 FIFO 中 (如果为 XYZ 配置了 MLC 滤波器, 则有三个字)。标记为 1Bh。
- MLC 标准特征**可在每个时间窗口(窗口长度由用户设置)存储到 FIFO 中。只有一个数据字存储到 FIFO 中, 标记为 1Ch。
- MLC 递归特征**可以在 MLC (MLC_ODR) 处理每个样本时存储到 FIFO 中。共有三个数据字存储到 FIFO 中, 标记是 1Ch。

表 12 列出了不同情况下的各个 FIFO 条目。滤波器和特征的标识符位于 Unico-GUI 生成的配置文件 (.ucf) 中 (参见图 20)。

表 12. FIFO 中的数据

	标记	X_L	X_H	Y_L	Y_H	Z_L	Z_H
MLC 结果	1Ah	MLCx_SRC	索引 (0,1,2,3)		时间戳		
MLC 滤波器	1Bh		值 ⁽¹⁾		滤波器标识符 ⁽²⁾	保留	
MLC 特征	1Ch		值 ⁽¹⁾		特征标识符 ⁽²⁾	保留	

1. 该值以半精度浮点值表示。

2. 滤波器和特征标识符在 .ucf 文件中指示。

图 20. 滤波器和特征的标识符

```
-- <MLC1_SRC>DT1,0='negative',4='positive'
-- FILTER_IIR1_ACC_X -> 023Ch
-- FILTER_IIR1_ACC_Y -> 023Eh
-- FILTER_IIR1_ACC_Z -> 0240h
-- F1_MEAN_on_ACC_X -> 0242h
-- F2_MEAN_on_ACC_Y -> 0244h
-- F3_MEAN_on_ACC_Z -> 0246h
```

下表列出了用于将结果和滤波器/特征批量存储到 FIFO 中的全局使能位。

位	寄存器	结果
MLC_FIFO_EN 位	EMB_FUNC_FIFO_EN_A (44h)	使能将机器学习内核结果批量存储到 FIFO 缓冲区中
MLC_FILTER_FEATURE_FIFO_EN 位	EMB_FUNC_FIFO_EN_B (45h)	使能将机器学习内核滤波器和特征批量存储到 FIFO 缓冲区中 (生成 MLC 配置时, 可以明确选择要存储到 FIFO 中的滤波器和特征。)

提示

必须在 MLC 配置期间通过 Unico-GUI 手动使能每一个滤波器和特征。

2 机器学习内核工具

通过专用工具可以实现对器件中的机器学习内核的可编程性，该专用工具可作为 Unico GUI 的扩展提供。

2.1 Unico GUI

Unico 是意法半导体产品组合中可用的所有 MEMS 传感器演示板的图形用户界面。它可以与基于 STM32 微控制器的主板（专业 MEMS 工具）进行交互，从而可以在 MEMS 传感器和 PC GUI 之间进行通信。Unico 也可以在未将主板连接到 PC 的情况下离线运行。

可在 www.st.com 上的 STEVAL-MKI109V3 中找到专业 MEMS 工具板的详细信息。

Unico GUI 提供支持三种操作系统（Windows、Linux、Mac OS X）的三个软件包。

Unico GUI 允许以图形或数字格式显示传感器输出，并允许用户保存或广泛管理来自器件的数据。

通过 Unico 可访问 MEMS 传感器寄存器，从而能够实现寄存器设置的快速原型设计，并可以直接在器件上轻松测试配置。可将当前寄存器配置保存在文本文件中（扩展名为 .ucf），并从现有文件中加载配置。这样，就可以在几秒内对传感器重新编程。

Unico GUI 中的机器学习内核工具通过自动生成器件的配置文件来抽象寄存器配置过程。用户只需在 GUI 中设置一些参数，然后点击几个按钮，便可以使用配置文件。用户可通过这些配置文件创建自己的器件配置库。

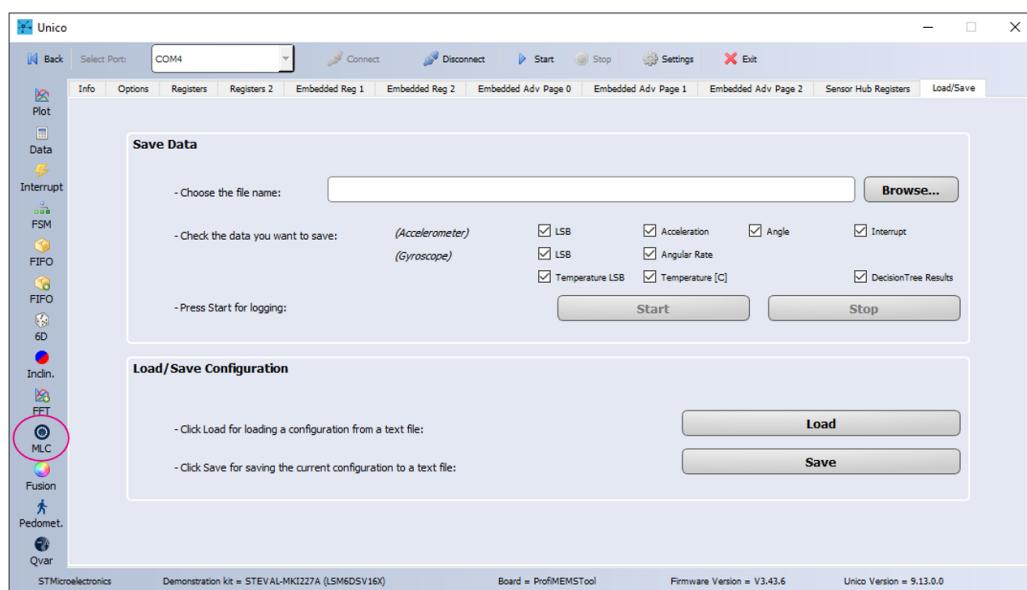
机器学习算法需要收集数据日志，可通过 Unico（图 21）的 [Load/Save]（加载/保存）选项卡获取数据日志。对于加速度计，[Acceleration]（加速度）复选框允许将数据保存为 [mg]。对于陀螺仪，[Angular rate]（角速率）复选框允许将数据保存为 [dps]。

提示

记录数据时，必须正确使用 [Start]（开始）和 [Stop]（停止）按钮（在 Unico 的 [Load/Save]（加载/保存）选项卡中），以免在采集开始或结束时记录错误的数据。例如，在记录“行走”类别的数据模式时，用户应在按下 [Start]（开始）按钮之前开始行走，并在按下 [Stop]（停止）按钮之后停止行走。必须选择正确的 ODR 来记录数据。如果已定义最终 MLC ODR（例如 30 Hz），建议使用该 ODR 来记录数据（ODR 30 Hz）。如果未定义 MLC ODR，建议以 ODR 120 Hz（即 MLC 的最大 ODR）记录数据，并在需要时对数据进行下采样。实现不同的算法时，需要的数据日志也会有所不同（每个类别至少需要一个数据日志，以便使用监督式机器学习方法）。建议每个类别使用多个不同的数据日志（例如，每个类别使用 30 个数据日志），以便捕获最终应用中可能存在的一些多样性或差异（例如，不同用户、不同测试或不同条件）。

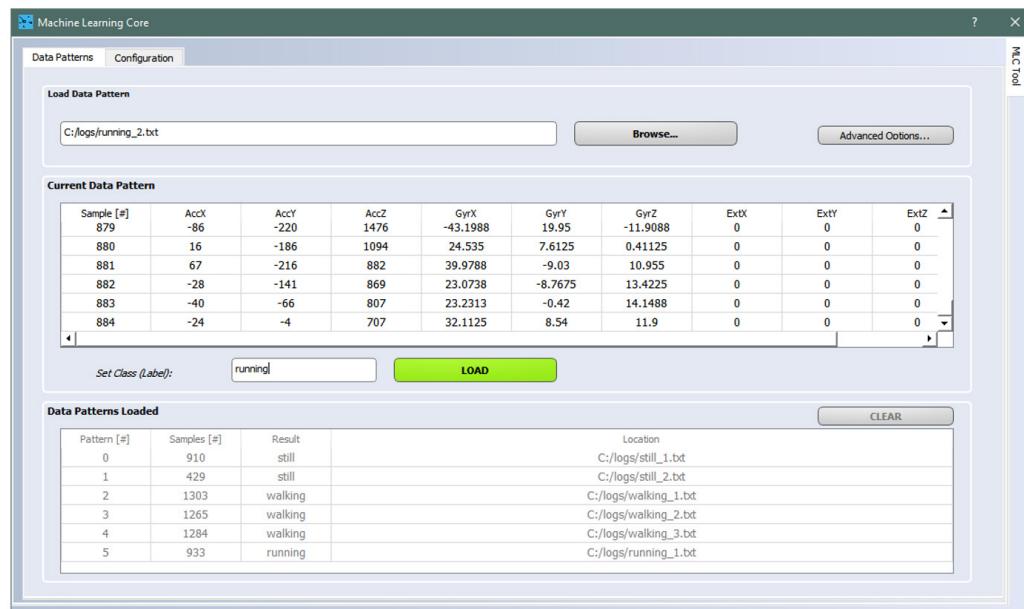
如果离线使用 Unico GUI（未将主板连接到 PC），则已经获取数据日志的用户可直接上传这些日志，以生成机器学习内核配置。

图 21. Unico GUI



然后可以使用 **[Data Patterns]** (数据模式) 选项卡 (图 22) 将收集到的数据日志载入 GUI 左侧的 Unico 机器学习内核工具中。必须将预期结果分配给每个加载的数据模式 (例如, 在活动识别算法中, 结果可能为: 静止、行走、慢跑等)。此分配也被称为“数据标记”。标记必须是一组只包含字母和数字的字符 (不包含特殊字符和空格)。此外, 还可以加载一组数据模式 (通过在文件所在文件夹中进行多次选择), 然后只为选定的所有文件分配一次标记。

图 22. 机器学习内核工具——数据模式



机器学习内核工具的 **[Data Patterns]** (数据模式) 选项卡中的预期数据测量单位为:

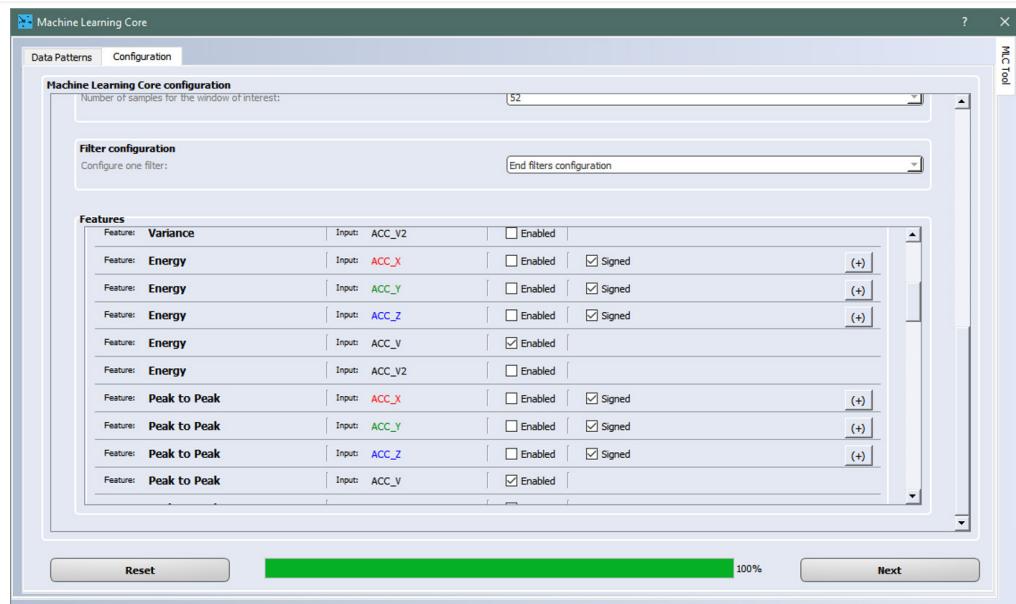
- 加速度计为 **[mg]** (或 **[g]**)
- 陀螺仪为 **[dps]** (或 **[mdps]**)

对于加速度计, 从 **[mg]** 转换为 **[g]**, 对于陀螺仪, 从 **[dps]** 转换为 **[rad/s]**, 并由机器学习内核工具从内部自动管理这些转换, 以使机器学习内核逻辑能够处理正确的数据 (**[g]** 和 **[rad/s]**)。对于外部传感器数据, 需要用户在配置的后期阶段设置适当的灵敏度。

在机器学习内核工具 (图 23) 的 **Configuration** (配置) 选项卡中, 可配置机器学习内核的所有参数 (如 ODR、满量程、窗口长度、滤波器、特征、元分类器)。该工具允许选择可应用于原始数据的多个滤波器, 并通过输入数据或滤波数据计算多个特征。计算出的特征将成为决策树的属性。

当已连接并配置好器件后, 该工具将根据当前的器件配置自动建议 ODR 和满量程(对于加速度计和陀螺仪)。

图 23. 机器学习内核工具——配置



机器学习内核工具的 **[Configuration]**（配置）选项卡将生成属性关系文件 (ARFF)，这是决策树生成过程的起点。决策树可由不同的机器学习工具生成（第 2.2 节）。

决策树生成后，可将其上传到 Unico 中的机器学习内核工具，以完成 LSM6DSV16X 的寄存器配置生成。

通过访问传感器寄存器，Unico GUI 可读取决策树输出状态，使其与传感器数据一起显示，并可以将所有数据（传感器输出和决策树输出）一起记录在同一文本文件中。

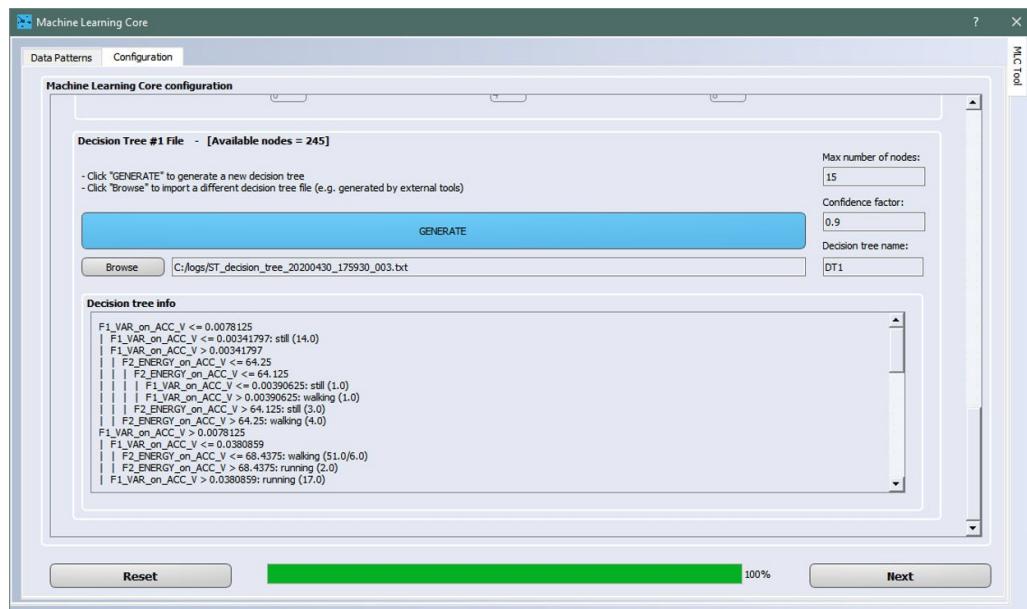
2.2 决策树生成

Unico（从版本 9.8 开始）可以自动生成决策树，如下图所示。

在 Unico 中启动决策树生成时，可以对以下两个参数进行微调：

- 最大节点数，通过设置该参数可以确保生成的决策树能够与 MLC 配置拟合；
- 置信系数，用于控制决策树剪枝（降低该值可减少过拟合）。

图 24. Unico 中的决策树生成



除了 Unico 之外，还有其他外部机器学习工具能够生成决策树，而且其中一些工具受 Unico 支持。

最常用的工具之一为怀卡托大学开发的 Weka 软件（有关此软件的更多详细信息，请参见附录 A）。其他替代工具包括：RapidMiner（附录 B）、MATLAB®（附录 C）、Python（附录 D）。

Weka 能够从属性关系文件 (ARFF) 生成决策树。通过 Weka 可以评估哪些属性适合决策树，并且可以通过更改 Weka 中的所有可用参数来实现不同的决策树配置。图 25 和图 26 显示了 Weka 的 [Preprocess]（预处理）和 [Classify]（分类）选项卡，这些选项卡可用于评估属性并生成决策树。

图 25. Weka Preprocess (预处理)

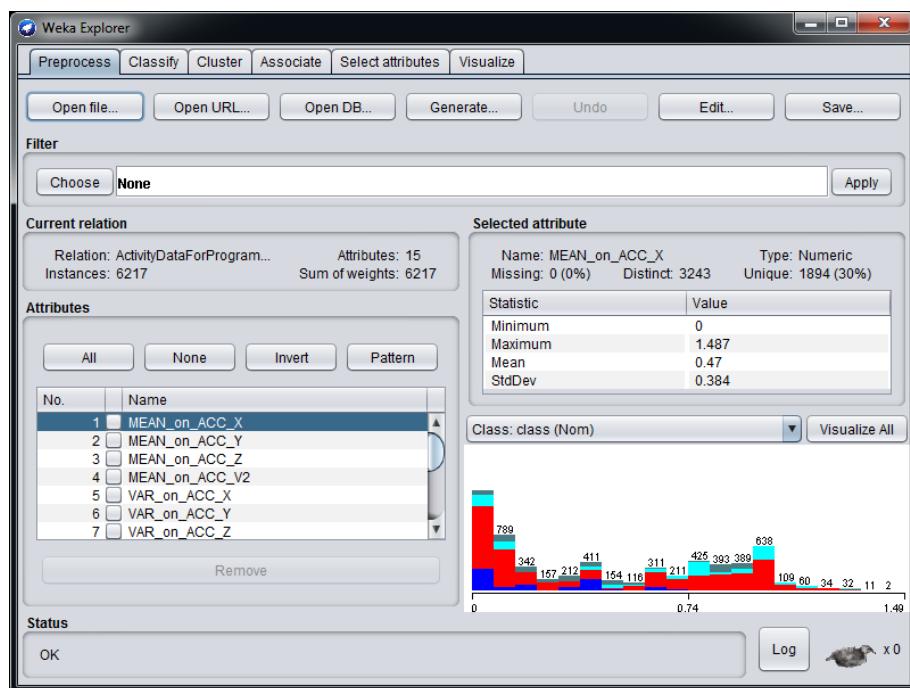
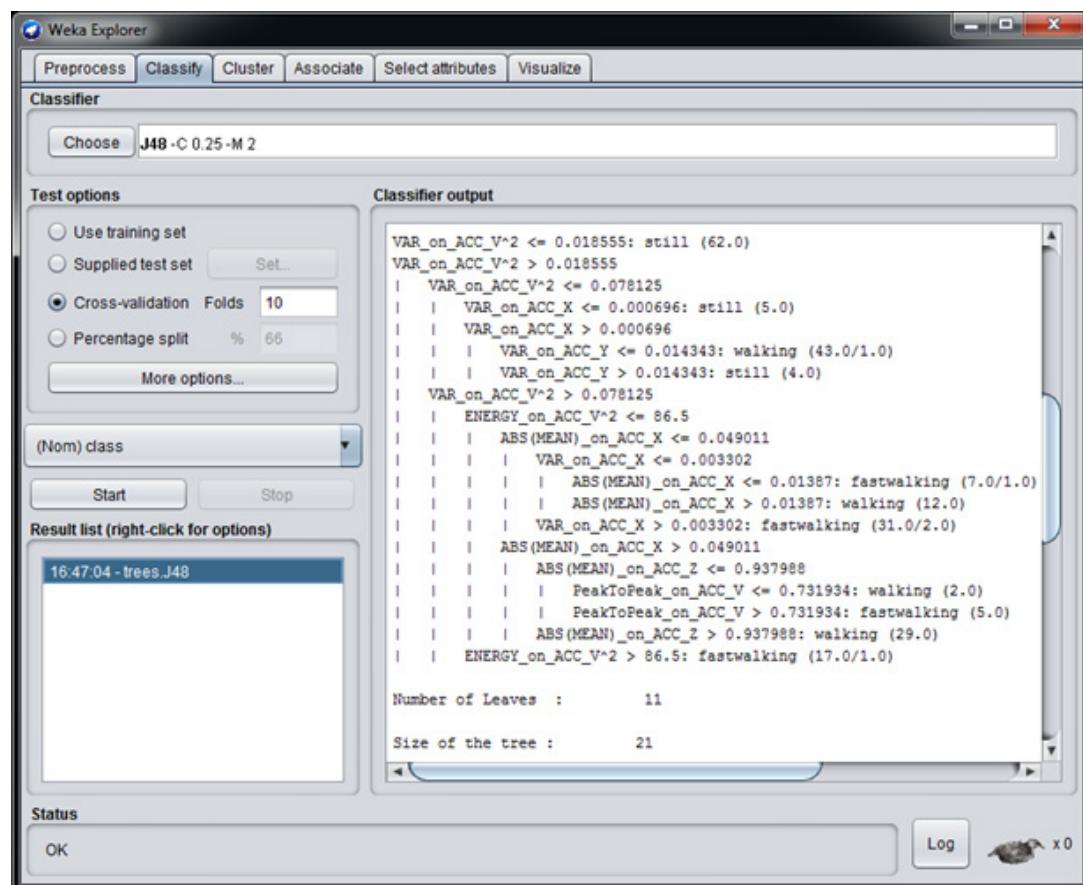


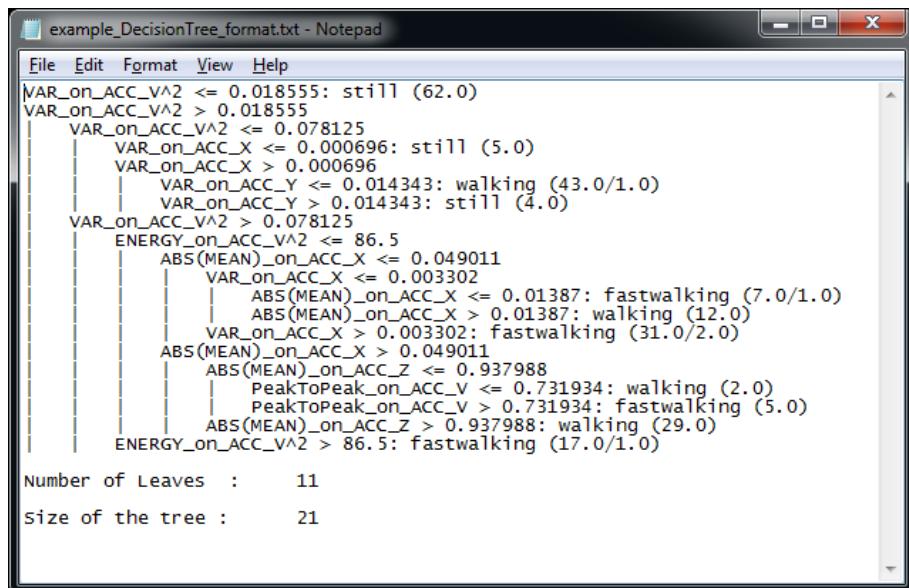
图 26. Weka Classify (分类)



决策树生成后，可将其上传到 Unico 中的机器学习内核工具，以完成 LSM6DSV16X 的寄存器配置生成。

Unico 中的机器学习内核工具允许以文本格式 (.txt) 输入决策树文件。文本文件必须包含采用 Weka J48 格式的决策树（决策树示例如图 27 中所示）。在 Weka [Classifier output]（分类器输出）（图 26）中，必须从第一行（第一个节点）选择决策树，或选择 RapidMiner 格式（附录 B）的决策树。最后两行（叶数和树的大小）可选。必须将 Weka 的选定输出复制到文本文件中。

图 27. 决策树格式



```
example_DecisionTree_format.txt - Notepad
File Edit Format View Help
VAR_on_ACC_V^2 <= 0.018555: still (62.0)
VAR_on_ACC_V^2 > 0.018555
| VAR_on_ACC_V^2 <= 0.078125
| | VAR_on_ACC_X <= 0.000696: still (5.0)
| | VAR_on_ACC_X > 0.000696
| | | VAR_on_ACC_Y <= 0.014343: walking (43.0/1.0)
| | | VAR_on_ACC_Y > 0.014343: still (4.0)
| | VAR_on_ACC_V^2 > 0.078125
| | ENERGY_on_ACC_V^2 <= 86.5
| | ABS(MEAN)_on_ACC_X <= 0.049011
| | | VAR_on_ACC_X <= 0.003302
| | | | ABS(MEAN)_on_ACC_X <= 0.01387: fastwalking (7.0/1.0)
| | | | ABS(MEAN)_on_ACC_X > 0.01387: walking (12.0)
| | | VAR_on_ACC_X > 0.003302: fastwalking (31.0/2.0)
| | | ABS(MEAN)_on_ACC_X > 0.049011
| | | ABS(MEAN)_on_ACC_Z <= 0.937988
| | | | PeakToPeak_on_ACC_V <= 0.731934: walking (2.0)
| | | | PeakToPeak_on_ACC_V > 0.731934: fastwalking (5.0)
| | | ABS(MEAN)_on_ACC_Z > 0.937988: walking (29.0)
| | ENERGY_on_ACC_V^2 > 86.5: fastwalking (17.0/1.0)

Number of Leaves : 11
Size of the tree : 21
```

如果决策树是由其他工具生成，则必须将格式转换为 Weka J48 格式（或 RapidMiner 格式），以便 Unico 中的机器学习内核工具能够正确读取决策树。

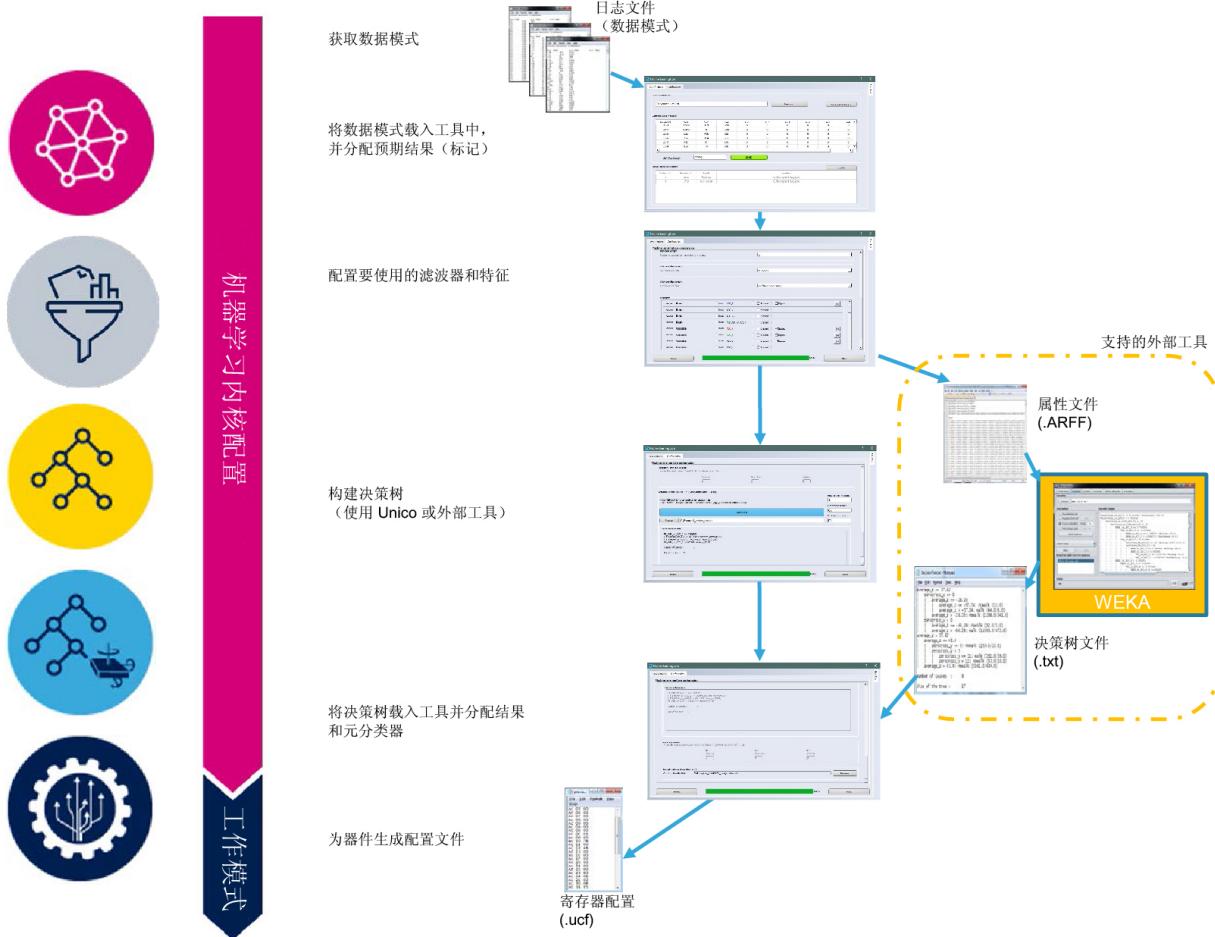
2.3 配置过程

图 28 显示从数据模式到生成器件 (LSM6DSV16X) 的寄存器设置的整个机器学习处理过程。

如第 2.1 节 Unico GUI 中所示，可通过 Unico GUI 的 [Load/Save]（加载/保存）选项卡获取数据模式。如果该操作无法实现，或如果用户希望使用一些不同的数据模式，用户仍可以将这些数据模式上传到 Unico 的机器学习内核工具中，但有一些限制：

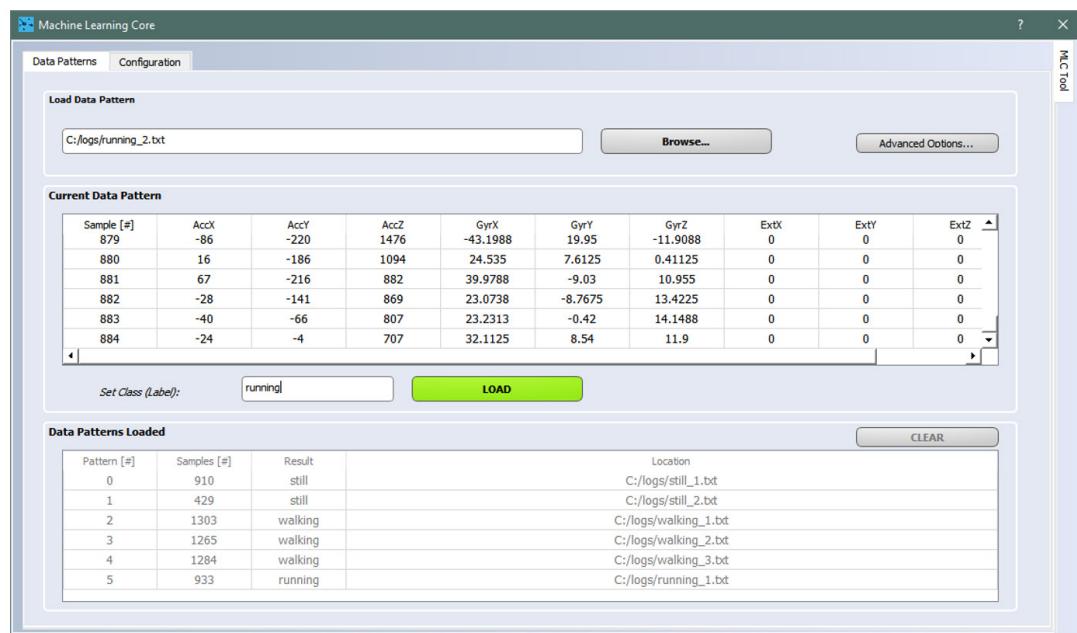
- 每种数据模式均必须以标题行开头，其中包含数据测量单位。
 - A_X [mg] A_Y [mg] A_Z [mg] G_X [dps] G_Y [dps] G_Z [dps]
- 标题行之后的数据必须用“制表符”或“空格”分隔。
- 文件列中的传感器顺序必须为加速度计数据（如果有）、陀螺仪数据（如果有）、外部传感器数据（如果有）。
- 任何传感器列中的轴顺序为 X、Y、Z。

图 28. 配置过程



打开 Unico 中的机器学习内核工具，可加载以上述格式获取的数据模式，并为每个数据日志分配预期结果（如下图所示）。

图 29. 将结果分配给数据模式



加载所有数据模式后，可通过 [Configuration]（配置）选项卡配置机器学习内核参数。这些参数为 ODR、满量程、决策树数量、窗口长度、滤波器、特征等（如图 30、图 31、图 32 和图 33 中所示）。

图 30. 机器学习内核配置

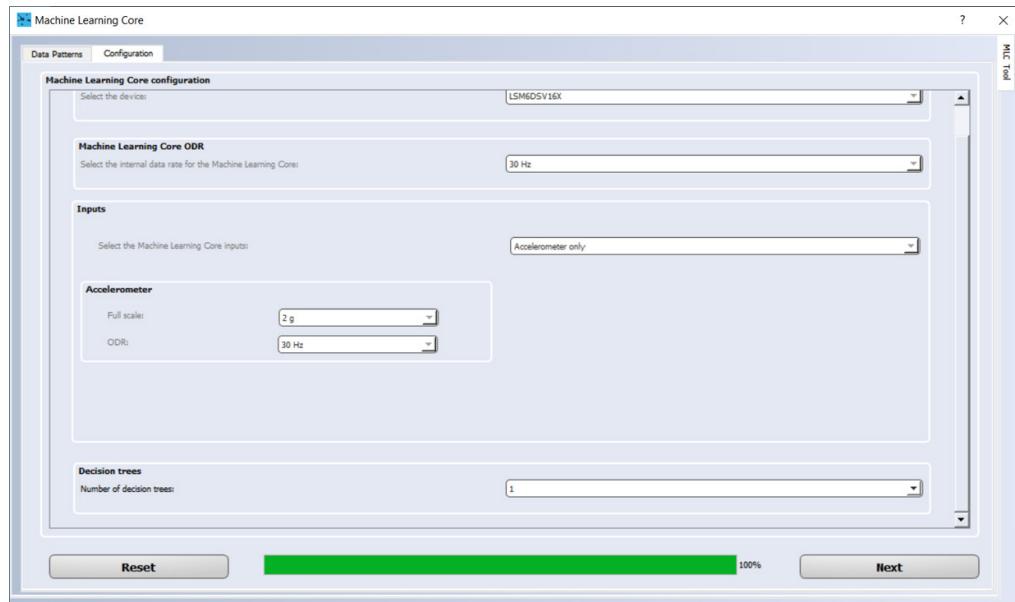


图 31. 滤波器配置

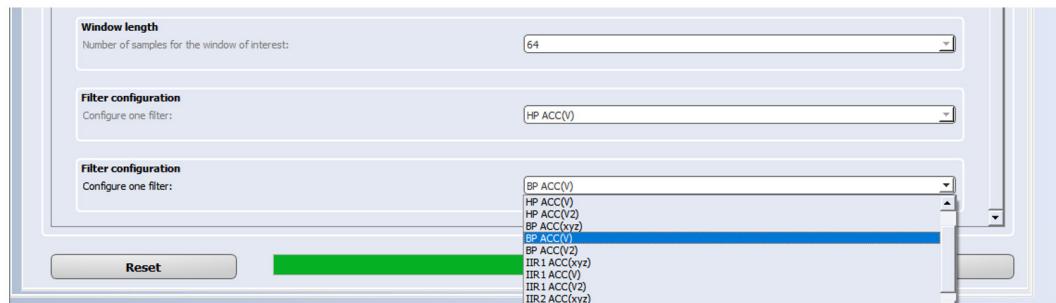
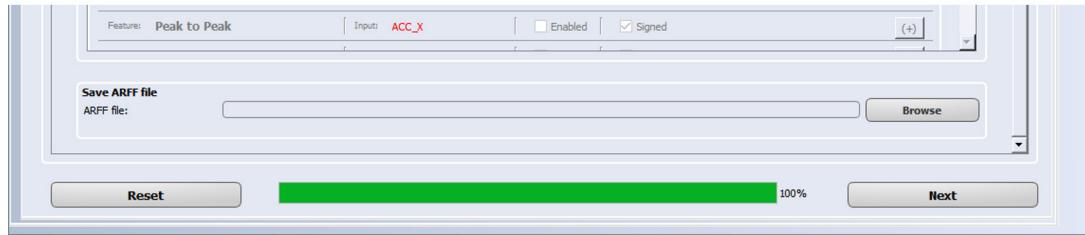


图 32. 特征配置



图 33. ARFF 生成



可选择多个滤波器和多个特征。在选择 **[End filter configuration]** (结束滤波器配置) 参数之前, GUI 会反复请求另一个滤波器 (图 31)。可使用复选框轻松选择所有可用特征 (图 32)。

配置完所有特征后，Unico 中的机器学习内核工具将生成一个 ARFF 文件（图 33），该文件包含根据训练数据计算出的所有特征。图 34 显示了 Unico 中的机器学习内核工具生成的 ARFF 文件示例。

Unico 内置决策树生成工具，该工具在内部使用 ARFF 文件来生成决策树，而 ARFF 文件也可以在外部工具（如 Weka）中使用。通过外部工具生成的决策树可以导入 Unico 中。

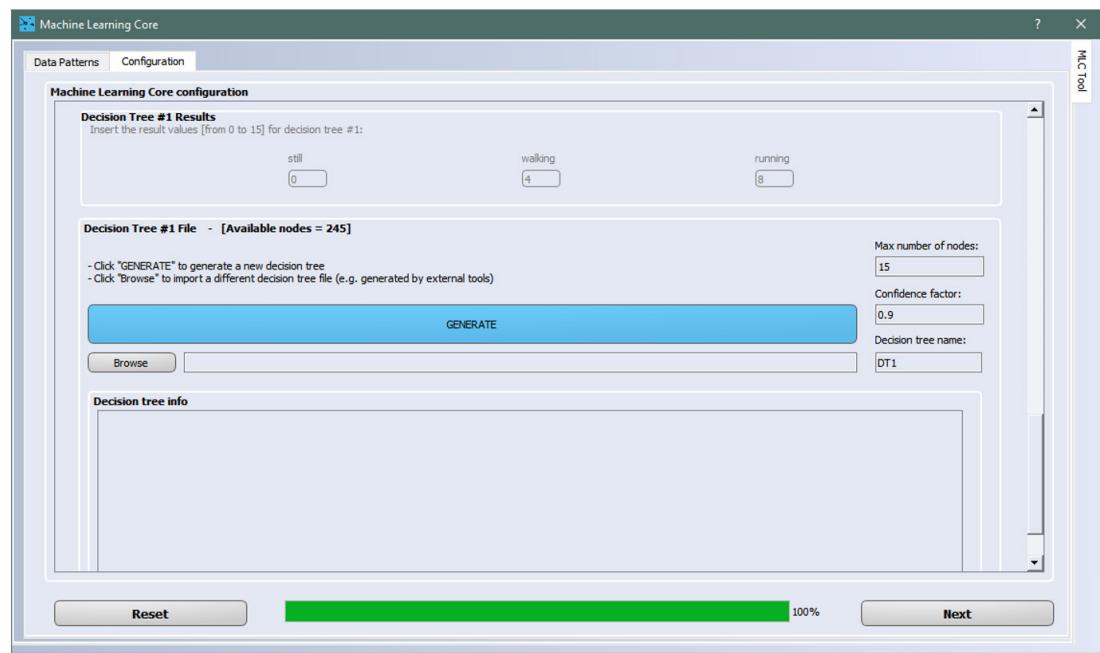
在某些情况下，用户必须将 ARFF 文件转换为其他外部决策树生成工具所需的文件格式，并且决策树必须与第 2.2 节 决策树生成中所述的格式兼容。有关支持的外部工具的更多信息，请参见附录部分。

图 34. ARFF 文件

```
C:\Vogel\example.aff - Notepad++ [Administrator]
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window
example.aff
1 @relation "MLG"
2
3 @attribute ABS(MEAN)_on_ACC_X numeric
4 @attribute ABS(MEAN)_on_ACC_Y numeric
5 @attribute ABS(MEAN)_on_ACC_Z numeric
6 @attribute ABS(MEAN)_on_ACC_V'2 numeric
7 @attribute ABS(VAR)_on_ACC_X numeric
8 @attribute ABS(VAR)_on_ACC_Y numeric
9 @attribute ABS(VAR)_on_ACC_Z numeric
10 @attribute ABS(VAR)_on_ACC_V'2 numeric
11 @attribute ABS(ENERGY)_on_ACC_X numeric
12 @attribute ABS(ENERGY)_on_filter_BP_on_ACC_V'2 numeric
13 @attribute ABS(ENERGY)_on_filter_IIR2_on_ACC_V'2 numeric
14 @attribute ABS(zeroCross)_on_filter_BP_on_ACC_V'2 numeric
15 @attribute ABS(zeroCross)_on_filter_IIR2_on_ACC_V'2 numeric
16 @attribute ABS(PeakToPeak)_on_ACC_Y numeric
17 @attribute class [Stationary, Walking, Jogging, Biking]
18
19 @data
20 0.0058403, 0.0390625, 0.11465, 1.03906, 3.8147e-06, 0.00683594, 0.0107422, 80.25, 0.585938, 73.75, 0, 1, 0.0146484, Stationary
21 0.0059967, 0.0389404, 0.11465, 0.103613, 2.02656e-06, 7.62939e-06, 0.00683594, 0.00878906, 79.875, 0.0004704, 79.9375, 0, 0, 0.0107422, Stationary
22 0.0064620, 0.0397987, 0.11465, 1.03233, 4.6916e-06, 4.76837e-06, 0.000976562, 0.00488281, 80.375, 0.000771523, 80.625, 0, 0, 0.0136719, Stationary
23 0.00595235, 0.0400696, 0.1127, 0.103516, 3.03984e-06, 7.62939e-06, 0.0078125, 0.00585938, 80, 0, 0.000483723, 80, 0, 0, 0.0126933, Stationary
24 0.0041624, 0.0390625, 1.03906, 0.103613, 1.0912e-06, 9.38374e-06, 0.000976562, 0.00488281, 80.175, 0.000936101, 80.0625, 0, 0, 0.0146484, Stationary
25 0.00595235, 0.0392765, 0.1166, 1.03516, 4.6916e-06, 4.76837e-06, 0.000976562, 0.00488281, 80.25, 0.000945689, 80.375, 0, 0, 0.0146484, Stationary
26 0.00514306, 0.0393677, 0.11465, 1.03125, 2.72752e-05, 7.62939e-06, 0.000976562, 0.00488281, 80, 0, 0, 0.0136719, Stationary
27 0.00475944, 0.0393677, 0.11465, 0.103613, 1.90755e-05, 1.14441e-05, 0.002929464, 0.002929464, 80.3125, 0.000469876, 80.5, 0, 0, 0.0107422, Stationary
28 0.00554267, 0.0393666, 0.11465, 1.03125, 3.5762e-05, 6.47572e-06, 0.000976562, 0.00488281, 80.05, 0, 0, 0.0107422, Stationary
29 0.00581743, 0.0394287, 0.11367, 0.103233, 1.61985e-05, 0.049404e-05, 0.002929464, 0.00195312, 80.125, 0.000677596, 80.5, 0, 0, 0.0126933, Stationary
30 0.00605392, 0.0397339, 0.1127, 0.103613, 2.62498e-05, 1.52034e-05, 0.000976562, 0.00683594, 80, 0, 0, 0.0107422, Stationary
31 0.00537109, 0.039093, 0.1127, 0.103613, 2.66102e-05, 6.67572e-06, 0.00878906, 0.00585938, 80.125, 0.000460425, 80.0425, 0, 0, 0.0136719, Stationary
32 0.00511613, 0.0394592, 0.11465, 1.03322, 2.26498e-05, 1.52858e-05, 0.002929464, 0.00195312, 80.25, 0, 0, 0.0117198, Stationary
33 0.00587463, 0.0387573, 0.11367, 0.103322, 3.63580e-06, 9.53674e-07, 0.00390625, 0.002929464, 80.3125, 0.00118351, 80.875, 0, 0, 0.0137581, Stationary
34 0.00461637, 0.0387594, 0.11367, 0.103613, 3.99351e-06, 7.62939e-06, 0.00488281, 0.00087906, 79.9375, 0.000247002, 80.875, 0, 0, 0.0107422, Stationary
35 0.00430569, 0.0393066, 0.11465, 0.103233, 3.15905e-06, 8.58337e-06, 0.000976562, 0.00390625, 80.25, 0.000857903, 80.1875, 0, 0, 0.0126933, Stationary
36 0.01744008, 0.0148163, 1.020501, 0.103613, 2.86102e-06, 0.00585938, 81.0625, 0.000492895, 81.4875, 0, 0, 0.0126933, Stationary
37 0.0185547, 0.014185, 0.101953, 0.040904, 1.86791e-06, 1.90735e-06, 0, 0.0078125, 80.625, 0.000478029, 81.125, 0, 0, 0.0117198, Stationary
38 0.0185549, 0.014185, 0.101953, 0.040904, 1.86791e-06, 1.90735e-06, 0, 0.0078125, 80.625, 0.000478029, 81.125, 0, 0, 0.0117198, Stationary
39 0.0181599, 0.014185, 0.101953, 0.040904, 1.86791e-06, 1.90735e-06, 0, 0.0078125, 80.625, 0.000478029, 81.125, 0, 0, 0.0117198, Stationary
40 0.0187370, 0.01435851, 0.101953, 0.040904, 5.48385e-06, 4.78837e-07, 0, 0.00625, 80.625, 0.000500779, 81.3625, 0, 0, 0.0117198, Stationary
41 0.0187370, 0.01435851, 0.101953, 0.040904, 4.78837e-06, 3.03994e-06, 0.000976562, 0.00488281, 80.875, 0.000270387, 81.3625, 0, 0, 0.00976562, Stationary
42 0.01824985, 0.0138016, 0.101855, 0.041029, 7.15256e-06, 2.26498e-04, 0.00390625, 0.00683594, 81, 0.000507832, 81.1875, 0, 0, 0.0146484, Stationary
43 0.01818141, 0.0145569, 0.102051, 0.103613, 6.43773e-06, 0.00488281, 0.000976562, 0.00683594, 81, 0.0625, 0.00332355, 81.875, 0, 0, 0.0137581, Stationary
44 0.01857501, 0.0141373, 0.103953, 0.040904, 5.94046e-06, 1.19209e-06, 0.00683594, 80.75, 0.000492895, 81.125, 0, 0, 0.0117198, Stationary
45 0.01721571, 0.0140076, 0.101953, 0.041029, 6.43773e-06, 2.96022e-06, 0.00292969, 0.00488281, 81.1875, 0, 0, 0.0136719, Stationary
Normal test file length: 4,215,755 lines: 31,380 Ln: Col:15 Sel:0|0 Windows (CR LF) UTF-8 IN5
```

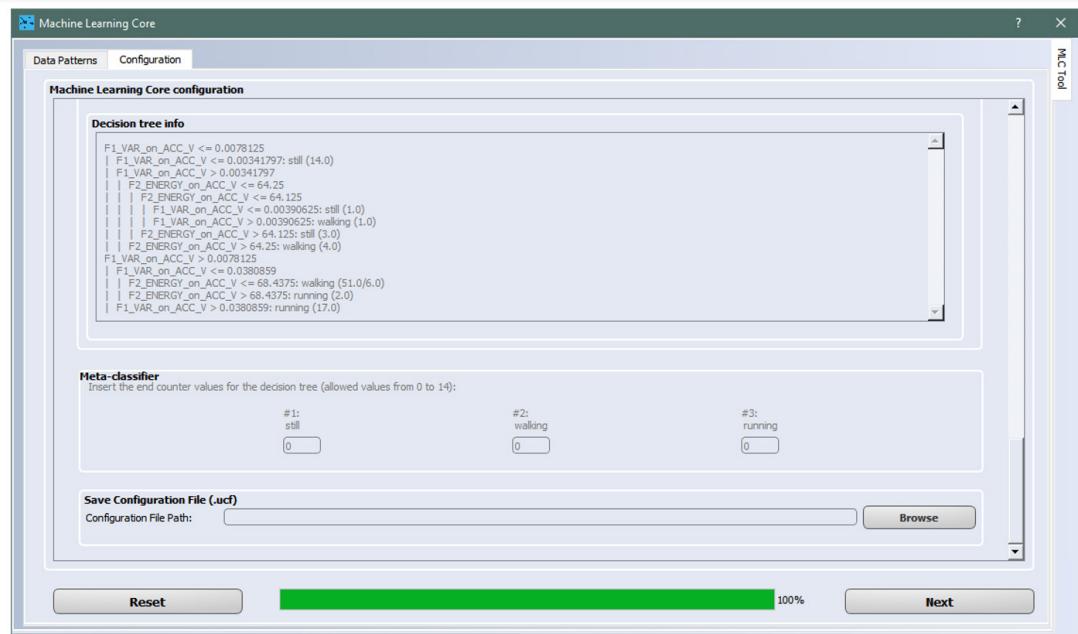
在生成或加载决策树之前，Unico 还会请求与决策树识别的每个类别相关的结果值。这些值用作 `MLCx_SRC` 寄存器的可能值。

图 35. 结果和决策树配置



配置过程的最后一步为配置 [元分类器]，这是用于生成决策树结果的可选滤波器。之后，工具已准备好生成器件的配置（图 36）。

图 36. 元分类器和器件配置



完成 MLC 配置后，便可在 Unico 中加载生成的 .ucf 文件，从而直接对器件进行编程。加载功能位于 Unico 主窗口的 [Load/Save]（加载/保存）选项卡中（图 38）。此外，在 MLC 配置结束后，可以通过勾选图 37 中所示的复选框直接加载器件上创建的配置。

图 37. 创建配置文件

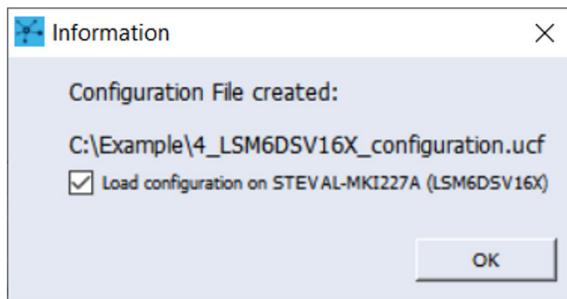


图 38. Unico 加载配置



对器件进行编程后，可在 Unico 的 [Data]（数据）窗口（图 39）中或在包含机器学习内核源寄存器（图 40）的其中一个寄存器选项卡中监控机器学习内核结果。

通过 Unico 生成的 .ucf 文件还可用于将生成的 MLC 配置集成到其他平台和软件中（如 AlgoBuilder、Unicleo、SensorTile.box 等）。

使用 .ucf 文件，还可以将值序列转换为头文件 (.h)，以便导入任何 C 项目（如驱动程序、固件等）：具体使用 Unico 主窗口的 options（选项）选项卡中的“C 代码生成”专用工具来生成 .h 文件（基于 .ucf 文件）。有关在标准 C 驱动程序中使用生成的 .h 文件的示例，请参见 GitHub 上的 [STMems_Standard_C_drivers 存储库]。

图 39. Unico 数据窗口

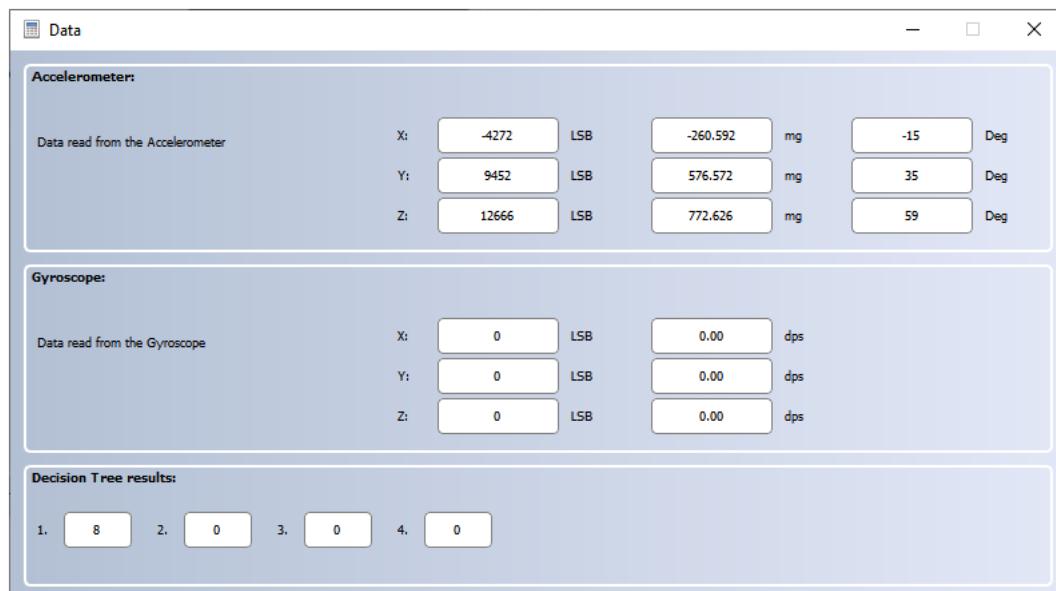


图 40. Unico——机器学习内核源寄存器

MLC1_SRC	(70h)	00	Read	Write	Default
MLC2_SRC	(71h)	00	Read	Write	Default
MLC3_SRC	(72h)	00	Read	Write	Default
MLC4_SRC	(73h)	00	Read	Write	Default

3 决策树示例

本节介绍了可载入 LSM6DSV16X 中的决策树的一些示例。

3.1 运动强度

下图中的决策树显示了仅使用加速度计范数中的“方差”特征所实现的运动强度的简单示例。该决策树可识别八种不同的强度级别。

本示例中所述的运动强度配置在低功耗模式下以 30 Hz 运行，并在具有 39 个样本的窗口中计算特征。LSM6DSV16X 在 1.8 V 下的电流消耗约为 12.6 μA 。关闭机器学习内核时，LSM6DSV16X 的电流消耗约为 10.5 μA ，因此机器学习内核的额外电流消耗仅为 2.1 μA 。

图 41. 运动强度决策树

```
dec_tree.txt - Notepad
File Edit Format View Help
module_variance <= 0.009: Intensity_0
module_variance > 0.009
|   module_variance <= 0.013671875: Intensity_1
|   module_variance > 0.013671875
|       module_variance <= 0.0234375: Intensity_2
|       module_variance > 0.0234375
|           module_variance <= 0.033203125: Intensity_3
|           module_variance > 0.033203125
|               module_variance <= 0.078125: Intensity_4
|               module_variance > 0.078125
|                   module_variance <= 0.1640625: Intensity_5
|                   module_variance > 0.1640625
|                       module_variance <= 0.3125: Intensity_6
|                       module_variance > 0.3125: Intensity_7

Number of Leaves : 8
size of the tree : 15
```

3.2 6D 位置识别

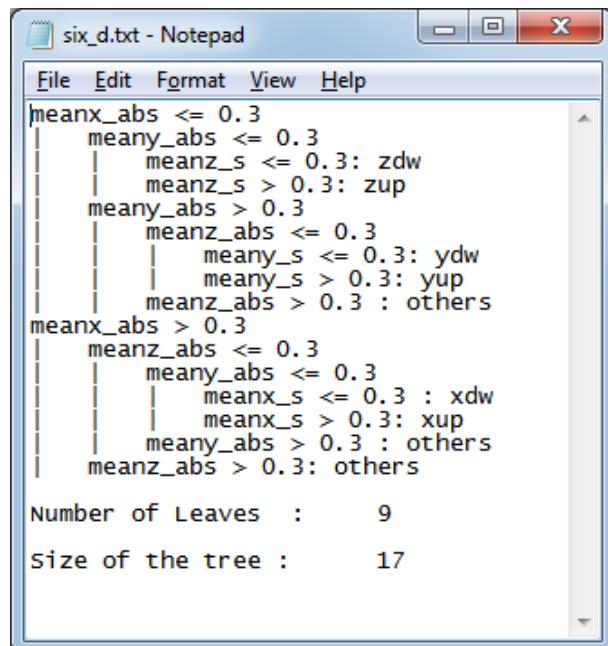
LSM6DSV16X 器件内置 6D 位置识别算法。本节中所述的示例仅显示了使用决策树的不同实现。

使用以下特征，可通过简单的决策树（图 42）轻松识别六种不同的位置：

- *meanx_abs*: 加速度计 X 轴的均值（无符号）
- *meany_abs*: 加速度计 Y 轴的均值（无符号）
- *meanz_abs*: 加速度计 Z 轴的均值（无符号）
- *meanx_s*: 加速度计 X 轴的均值（带符号）
- *meany_s*: 加速度计 Y 轴的均值（带符号）
- *meanz_s*: 加速度计 Z 轴的均值（带符号）

本示例中所述的 6D 位置识别配置在低功耗模式下以 30 Hz 运行，并在具有 20 个样本的窗口中计算特征。LSM6DSV16X 在 1.8 V 下的电流消耗约为 13.7 μA 。关闭机器学习内核时，LSM6DSV16X 的电流消耗约为 10.5 μA ，因此机器学习内核的额外电流消耗仅为 3.2 μA 。

图 42. 6D 决策树



The screenshot shows a Windows Notepad window titled "six_d.txt - Notepad". The content displays a decision tree structure with the following rules:

```
meanx_abs <= 0.3
|   meany_abs <= 0.3
|   |   meanz_s <= 0.3: zdw
|   |   meanz_s > 0.3: yup
|   meany_abs > 0.3
|   |   meanz_abs <= 0.3
|   |   |   meany_s <= 0.3: ydw
|   |   |   meany_s > 0.3: yup
|   |   meanz_abs > 0.3 : others
meanx_abs > 0.3
|   meanz_abs <= 0.3
|   |   meany_abs <= 0.3
|   |   |   meanx_s <= 0.3 : xdw
|   |   |   meanx_s > 0.3: xup
|   |   meany_abs > 0.3 : others
|   meanz_abs > 0.3: others

Number of Leaves :      9
Size of the tree :     17
```

3.3 健身活动识别

健身活动识别旨在作为可穿戴设备（如智能手表或腕带）的健身示例。为通过决策树实现该算法，已使用安装在腕带上的器件 (LSM6DSV16X) 获取了所有数据日志。

使用了加速度计输入（数据速率为 30 Hz），并在具有 90 个样本的窗口中计算了不同的特征，如表 13 所示。图 43 中的决策树可用于识别不同的健身活动，包括屈臂、侧平举和深蹲。

本示例中所述的健身活动识别配置以 30 Hz 运行，并在具有 90 个样本的窗口中计算特征。LSM6DSV16X 在 1.8 V 下的电流消耗约为 17 μ A。关闭机器学习内核时，LSM6DSV16X（加速度计在低功耗模式下以 30 Hz 运行）的电流消耗约为 10.5 μ A，因此该算法的机器学习内核的额外电流消耗为 6.5 μ A。

表 13. 健身活动识别配置

配置	加速度计, 30 Hz ODR, 低功耗模式, 8 g 满量程
窗口长度	90 个样本（约 3 秒）
滤波器	加速度计 X/Y/Z 上为 IIR1
特征	均值、峰峰值、最小值、最大值
输出	无活动 (0)
	屈臂 (4)
	侧平举 (8)
	深蹲 (12)
元分类器	所有分类均为 0

图 43. 健身活动识别决策树

```

gym_activity_recognition.txt - Notepad
File Edit Format View Help
F5_MEAN_on_filter_IIR1_on_ACC_Z <= -0.1
|   F1_MEAN_on_filter_IIR1_on_ACC_X <= 0.48
|   |   F3_MEAN_on_filter_IIR1_on_ACC_Y <= 0.2
|   |   |   F10_MIN_on_filter_IIR1_on_ACC_X <= -0.5
|   |   |   |   F13_MAX_on_filter_IIR1_on_ACC_X <= 0.2: nothing
|   |   |   |   F13_MAX_on_filter_IIR1_on_ACC_X > 0.2
|   |   |   |   |   F15_MAX_on_filter_IIR1_on_ACC_Z <= 0.7
|   |   |   |   |   |   F12_MIN_on_filter_IIR1_on_ACC_Z <= -0.5
|   |   |   |   |   |   |   F8_PeakToPeak_on_filter_IIR1_on_ACC_Y <= 1.0: biceps
|   |   |   |   |   |   |   F8_PeakToPeak_on_filter_IIR1_on_ACC_Y > 1.0: nothing
|   |   |   |   |   |   |   F12_MIN_on_filter_IIR1_on_ACC_Z > -0.5: nothing
|   |   |   |   |   |   |   F15_MAX_on_filter_IIR1_on_ACC_Z > 0.7: nothing
|   |   |   |   |   |   F10_MIN_on_filter_IIR1_on_ACC_X > -0.5: nothing
|   |   |   |   F3_MEAN_on_filter_IIR1_on_ACC_Y > 0.2: nothing
|   |   |   F1_MEAN_on_filter_IIR1_on_ACC_X > 0.48: nothing
F5_MEAN_on_filter_IIR1_on_ACC_Z > -0.1
|   F5_MEAN_on_filter_IIR1_on_ACC_Z <= 0.35: nothing
|   F5_MEAN_on_filter_IIR1_on_ACC_Z > 0.35
|   |   F5_MEAN_on_filter_IIR1_on_ACC_Z <= 0.85
|   |   |   F1_MEAN_on_filter_IIR1_on_ACC_X <= -0.2
|   |   |   |   F4_ABS_MEAN_on_filter_IIR1_on_ACC_Y <= 0.6
|   |   |   |   |   F13_MAX_on_filter_IIR1_on_ACC_X <= 0.6
|   |   |   |   |   |   F13_MAX_on_filter_IIR1_on_ACC_X <= -0.3: nothing
|   |   |   |   |   |   F13_MAX_on_filter_IIR1_on_ACC_X > -0.3
|   |   |   |   |   |   |   F10_MIN_on_filter_IIR1_on_ACC_X <= -0.6
|   |   |   |   |   |   |   |   F10_peakToPeak_on_filter_IIR1_on_ACC_Y <= 1.0: nothing

```

3.4 示例汇总

下表汇总显示了本文档中所述的所有示例，以显示 LSM6DSV16X 中的机器学习内核在不同配置下的典型电流消耗。主要贡献者为机器学习内核 ODR（可能与传感器 ODR 不同）、配置的决策树数量以及每个决策树的节点数。

表 14. 示例汇总

算法	MLC_ODR	决策树数量	节点数	MLC 额外电流消耗
运动强度	30 Hz	1	7	2.1 μ A
6D 位置识别	30 Hz	1	8	3.2 μ A
健身活动识别	30 Hz	1	28	6.5 μ A

附录 A Weka

Weka 是由新西兰的怀卡托大学开发的免费软件。它包含一系列用于数据分析和预测建模的可视化工具与算法，以及用于轻松访问这些功能的图形用户界面。

Weka 是最受欢迎的决策树生成机器学习工具之一。本节包含有关该外部软件的一些详细信息，可在以下链接中找到其他详细信息：

- [Weka 下载](#)
- [Weka 网站](#)
- [Weka 用户指南](#)

所有 Weka 技术均基于以下假设，即数据可作为一个平面文件或关系来使用，并通过固定数量的属性来描述每个数据点。

ARFF（属性关系文件格式）文件是 ASCII 文本文件，描述了共享一组属性的实例列表。ARFF 文件具有两个不同的部分，如图 44 中所示：包含属性（特征、类别）的头部分，包含所有特征值以及与该特征集相关的相应类别的数据部分。

图 44. ARFF 示例

```
C:\Vogs\example.arff - Notepad++ [Administrator]
File Edt Search View Encoding Language Settings Tgols Macro Run Plugins Window
example.arff
1  @relation 'MLG'
2
3  @attribute ABS(MEAN)_on_ACC_X numeric
4  @attribute ABS(MEAN)_on_ACC_Y numeric
5  @attribute ABS(MEAN)_on_ACC_Z numeric
6  @attribute ABS(VAR)_on_ACC_X numeric
7  @attribute ABS(VAR)_on_ACC_Y numeric
8  @attribute ABS(VAR)_on_ACC_Z numeric
9  @attribute ABS(ENERGY)_on_ACC_V2 numeric
10 @attribute ABS(ENERGY)_on_filter_BP_on_ACC_V2 numeric
11 @attribute ABS(ENERGY)_on_filter_IIR2_on_ACC_V2 numeric
12 @attribute ABS(zeroCross)_on_filter_BP_on_ACC_V2 numeric
13 @attribute ABS(zeroCross)_on_filter_IIR2_on_ACC_V2 numeric
14 @attribute ABS(PeakToPeak)_on_ACC_V numeric
15 @attribute ABS(PeakToPeak)_on_ACC_V2 numeric
16 @attribute class {Stationary, Walking, Jogging, Biking}
17
18
19 #data
20 0.0058403, 0.0390625, 1.01465, 1.03906, 4.17233e-06, 3.8147e-06, 0.00683594, 0.0107422, 80.25, 0.585938, 73.75, 0, 1, 0.0146484, Stationary
21 0.0059367, 0.039404, 1.01465, 1.03613, 2.02656e-06, 7.62933e-06, 0.00683594, 0.0087806, 79.875, 0.0004704, 79.9375, 0, 0, 0.0107422, Stationary
22 0.004099207, 0.039978, 1.01465, 1.03223, 4.4916e-06, 4.76837e-06, 0.00976562, 0.004858201, 80.375, 0.00071523, 80.625, 0, 0, 0.0136719, Stationary
23 0.00559235, 0.0400696, 1.0127, 1.03516, 3.03984e-06, 7.62933e-06, 0.0078125, 0.00585938, 80, 0.00043273, 80, 0, 0, 0.0126953, Stationary
24 0.00618744, 0.0392151, 1.01562, 1.03018, 3.45707e-06, 8.83807e-06, 0.000976562, 0.000426821, 80.025, 0.000636101, 80.0625, 0, 0, 0.0146484, Stationary
25 0.00530624, 0.0394952, 1.01466, 1.0332, 4.03312e-06, 9.53674e-06, 0.00195312, 0.000976562, 80.1875, 0.000390053, 80.125, 0, 0, 0.0126953, Stationary
26 0.00593127, 0.0392761, 1.01466, 1.03018, 3.69549e-06, 7.62933e-06, 0.00676562, 0.000426821, 80.075, 0.00070549, 80.9375, 0, 0, 0.0126953, Stationary
27 0.00584276, 0.0393666, 1.01466, 1.03018, 3.69549e-06, 7.62933e-06, 0.00676562, 0.000426821, 80.075, 0.00070549, 80.9375, 0, 0, 0.0126953, Stationary
28 0.00675964, 0.0393999, 1.01562, 1.03013, 1.90755e-06, 1.14441e-06, 6.67572e-06, 0.002929469, 80.3125, 0.000609875, 80.5, 0, 0, 0.0107422, Stationary
29 0.00554276, 0.0393566, 1.01465, 1.03125, 3.57627e-06, 8.67572e-06, 0.000976562, 0.000976562, 80.075, 0.00049725, 80.125, 0, 0, 0.0107422, Stationary
30 0.00581741, 0.0394287, 1.01347, 1.03223, 6.19852e-06, 1.04904e-05, 0.002929469, 0.00195312, 80.125, 0.000477586, 80.5, 0, 0, 0.0126953, Stationary
31 0.00405392, 0.0397339, 1.0127, 1.03613, 2.5034e-05, 1.04904e-05, 0.00976562, 0.00683594, 80.0625, 0.00049725, 80, 0, 0, 0.0107422, Stationary
32 0.00537109, 0.039093, 1.0127, 1.03613, 2.86102e-05, 6.67572e-06, 0.00878906, 0.00585938, 80.125, 0.000460625, 80.0425, 0, 0, 0.0136719, Stationary
33 0.004651169, 0.0394952, 1.01466, 1.03302, 2.26498e-06, 1.52588e-05, 0.00292969, 0.00195312, 80.25, 0.000424862, 80.25, 0, 0, 0.0117188, Stationary
34 0.00587463, 0.0397573, 1.01367, 1.03302, 3.63588e-06, 9.53674e-07, 0.00390625, 0.002929469, 80.3125, 0.00183381, 80.875, 0, 0, 0.0175781, Stationary
35 0.00416837, 0.0398794, 1.01367, 1.03013, 3.99351e-06, 7.62933e-06, 0.00488281, 0.0087896, 79.9375, 0.000247002, 80.875, 0, 0, 0.0107422, Stationary
36 0.00584276, 0.0393666, 1.01466, 1.03013, 3.69549e-06, 7.62933e-06, 0.00676562, 0.000426821, 80.075, 0.00070549, 80.9375, 0, 0, 0.0126953, Stationary
37 0.0174405, 0.0148161, 1.03951, 1.04114, 6.43736e-06, 2.46102e-06, 0.00585938, 81.0625, 0.000485095, 81.875, 0, 0, 0.0126953, Stationary
38 0.0185547, 0.0141818, 1.01985, 1.04004, 7.46781e-06, 0, 0.0078125, 80.425, 0.000475028, 81.125, 0, 0, 0.0117188, Stationary
39 0.0185549, 0.0135422, 1.01855, 1.03711, 5.72205e-06, 4.76837e-07, 0, 0.00878906, 81.0425, 0, 0, 0.0126953, Stationary
40 0.0183105, 0.0145569, 1.01953, 1.03906, 5.49363e-06, 1.66883e-06, 0.00976562, 0.002929469, 80.875, 0, 0.000500479, 81.5625, 0, 0, 0.0117188, Stationary
41 0.0187378, 0.0140585, 1.01953, 1.04004, 6.43736e-06, 3.09944e-06, 0.000976562, 0.00683594, 80.6875, 0, 0.000270367, 81.1875, 0, 0, 0.00976562, Stationary
42 0.0182495, 0.0138016, 1.01855, 1.04199, 7.15256e-06, 2.26498e-06, 0.00390625, 0.00683594, 81, 0.000507832, 81.1875, 0, 0, 0.0146484, Stationary
43 0.0188141, 0.0145569, 1.02051, 1.03613, 6.43736e-06, 2.26498e-06, 0.00488281, 0.00683594, 81.0625, 0.000332355, 81.875, 0, 0, 0.0117188, Stationary
44 0.0187531, 0.0141373, 1.01953, 1.04004, 5.96046e-06, 1.19209e-06, 0, 0.00683594, 80.75, 0.000498295, 81.125, 0, 0, 0.0117188, Stationary
45 0.0175171, 0.0140076, 1.01953, 1.04199, 6.43736e-06, 2.98023e-06, 0.002929469, 80.0488281, 81.1875, 0.0011692, 81.75, 0, 0, 0.0136719, Stationary
```

图 45. Weka GUI Chooser (Weka GUI 选择器)



启动 Weka 时，将显示 [Weka GUI Chooser] (Weka GUI 选择器) 窗口（图 45），可通过第一个按钮选择的 [Explorer] (浏览器) 部分是 Weka 的主用户界面。

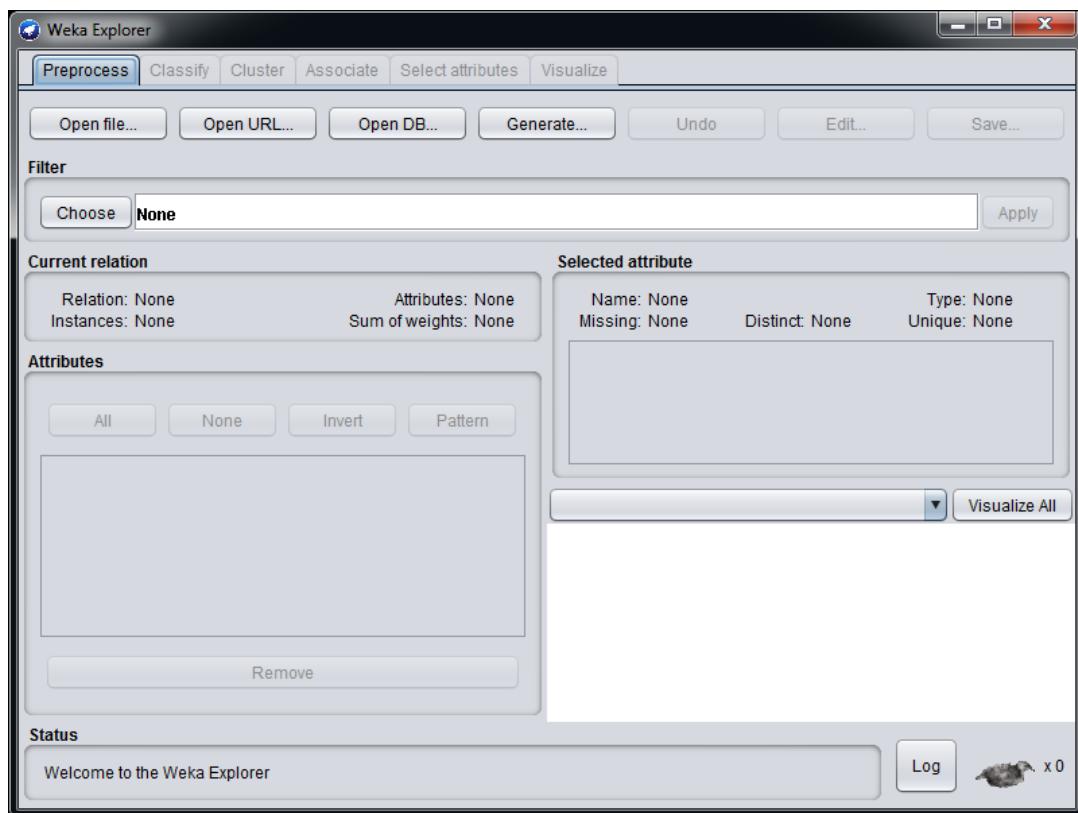
选择 Weka [Explorer] (浏览器)，将出现一个新界面（图 46）。[Explorer] (浏览器) 界面中有几个选项卡：

- [Preprocess] (预处理) 选项卡具有导入数据的工具。
- [Classify] (分类) 选项卡允许将分类和回归算法应用于数据集，以估计所产生的预测模型的精度并显示错误预测。
- 通过 [Cluster] (群集) 选项卡可访问 Weka 中的群集技术。
- 通过 [Associate] (关联) 选项卡可访问关联规则学习者，这些学习者试图识别数据中属性之间的所有重要的相互关系。
- [Select attributes] (选择属性) 选项卡提供了用于识别数据集中最具预测性的属性的算法。
- [Visualize] (可视化) 选项卡显示了散点图矩阵。

该附录部分仅描述了 [Preprocess] (预处理) 选项卡和 [Classify] (分类) 选项卡。

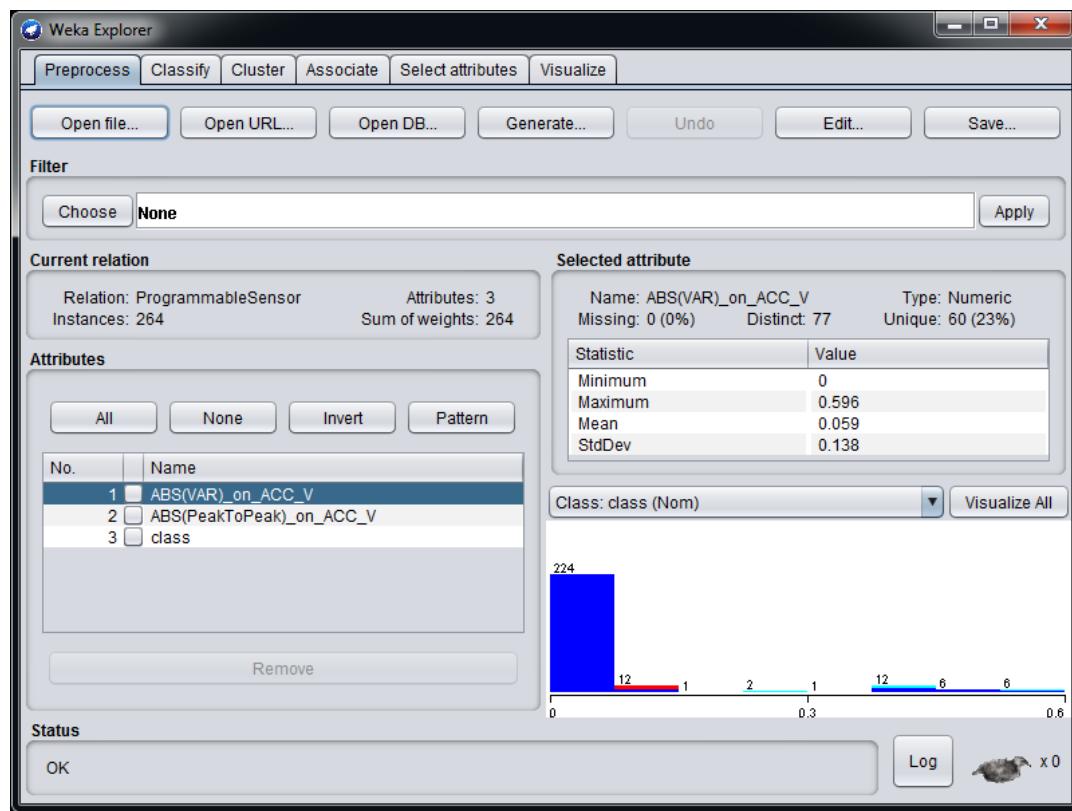
[Preprocess] (预处理) 选项卡如图 46 所示，它允许通过 [Open file] (打开文件) 按钮加载 ARFF 文件。

图 46. Weka Explorer (浏览器)



加载 ARFF 文件后, [Preprocess] (预处理) 选项卡将显示导入的 ARFF 文件的所有属性 (功能和类别)。属性可以图形化方式显示, 用户可以选择用于分类的属性。

图 47. Weka Explorer (浏览器) —— 属性



选择属性后，可在 Weka [Explorer] (浏览器) 的 [Classify] (分类) 选项卡中配置分类器 (图 48)。Weka 中有许多可用的分类器：通过选择分类器 [J48] (在 [trees] (树) 下)，可以生成决策树 (图 49)。

图 48. Weka 分类

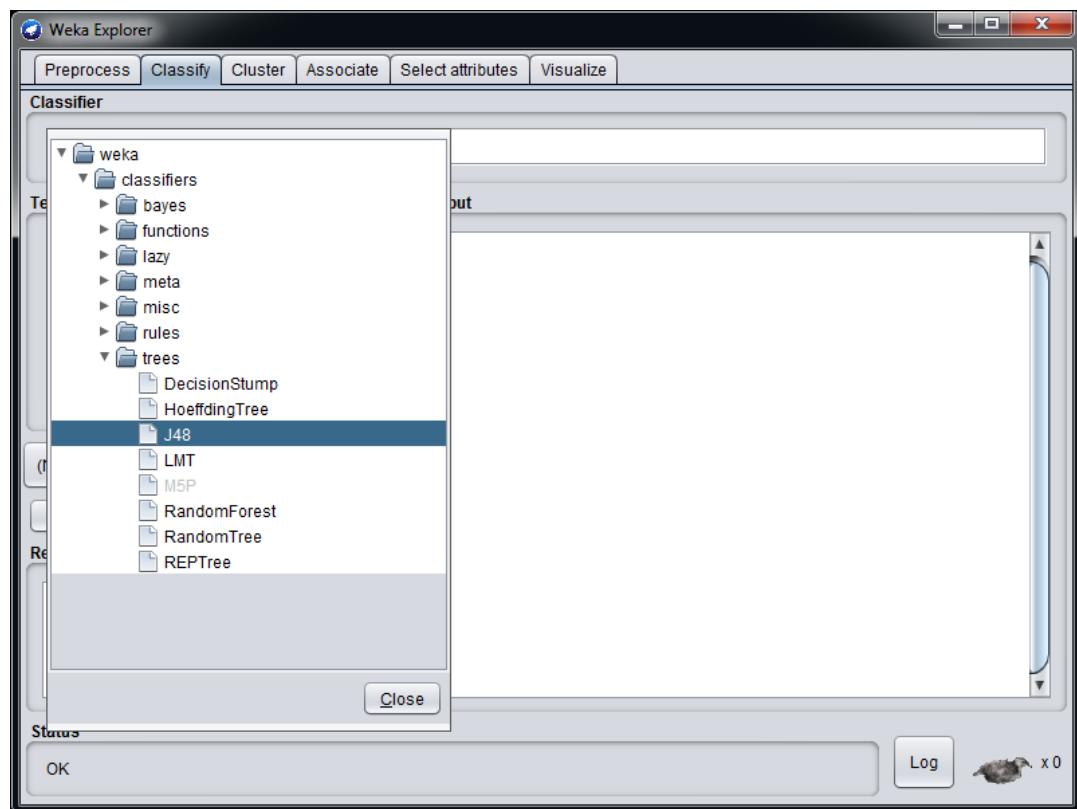
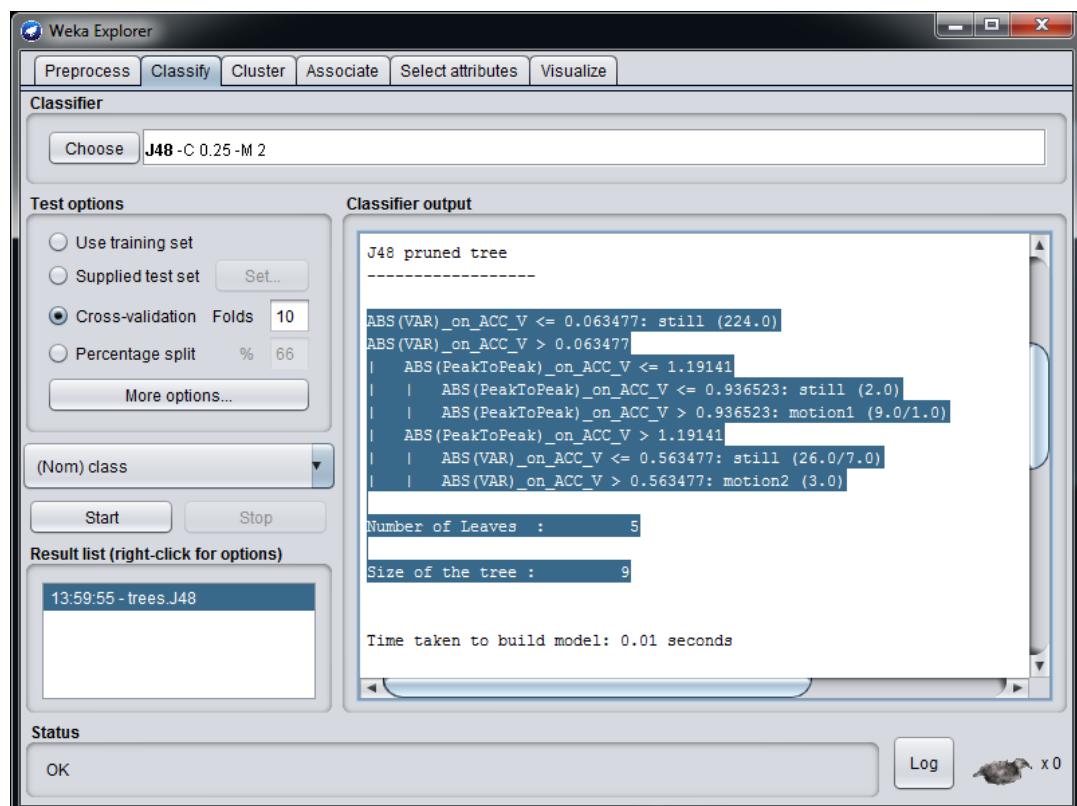
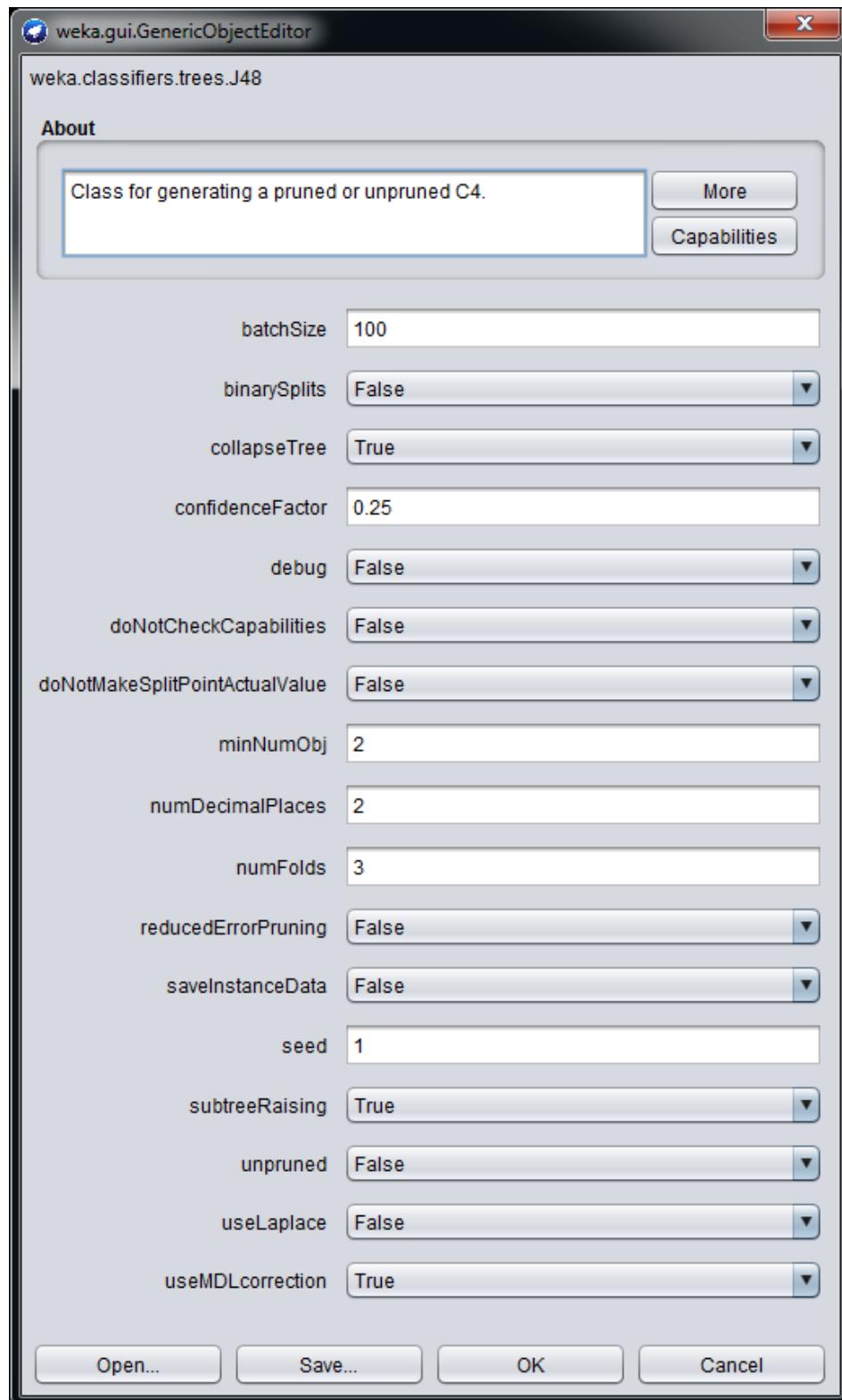


图 49. Weka 分类 J48



可更改分类器部分中的许多参数(图 50)，通过单击 [Start] (开始) 按钮，可生成不同的决策树(参见图 49)。

图 50. Weka J48 分类器参数



可轻松比较所有生成的决策树的以下方面：

- 节点数

由于 Weka 中的 J48 算法生成的决策树为二叉树，因此可通过从决策树之后的第一行出现的“叶数”参数减 1 来获得节点数（参见图 51. 正确分类实例）。此外，还可右键单击工具左侧的 [Result list]（结果列表）部分（在此可以轻松比较已创建的所有模型），以图形方式显示决策树。

- 正确分类实例

这是对所创建模型的精度的估计。将模型的结果与标签提供的结果进行比较。图 51. 正确分类实例显示了活动识别模型的正确分类实例。

- 混淆矩阵

NxN 表总结了分类模型预测的成功程度，即标签与模型分类的相关性。混淆矩阵的一个轴为预测模型标签，另一个轴为实际标签。

图 51. 正确分类实例

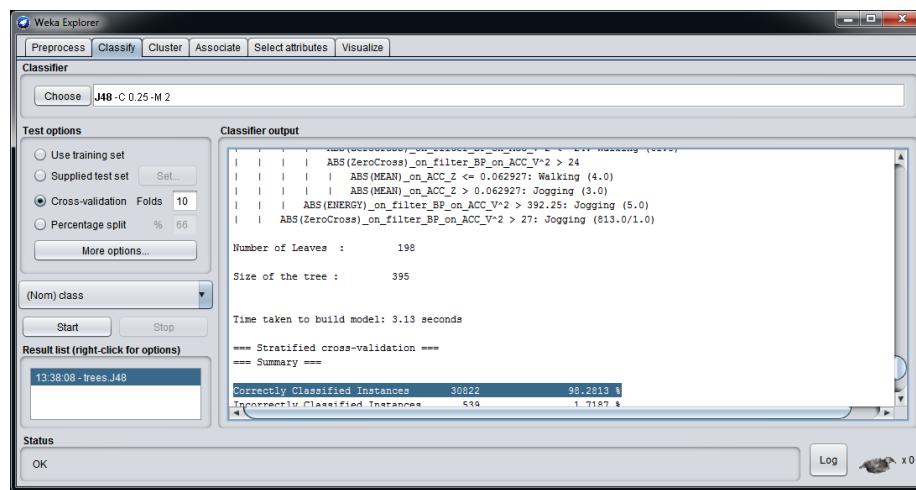
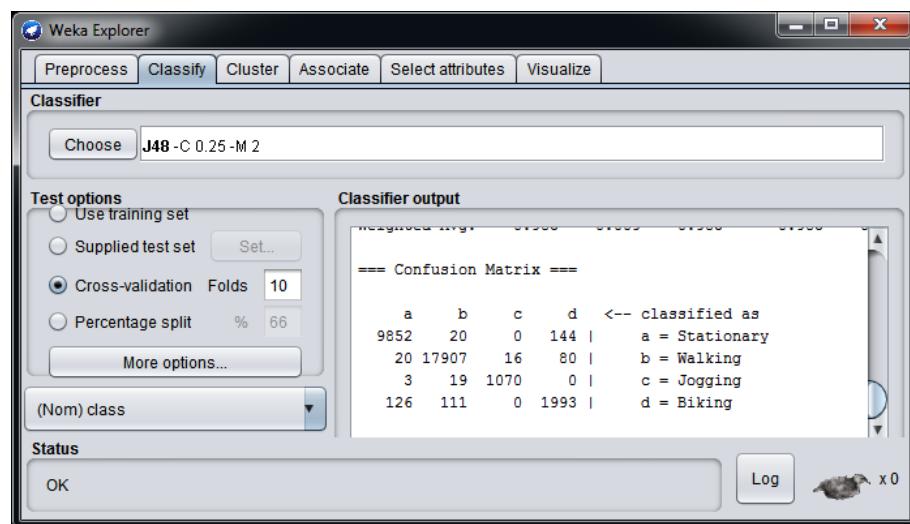


图 52. 混淆矩阵显示了具有四个类别（静止、行走、慢跑、骑自行车）的活动识别算法的混淆矩阵示例。

图 52. 混淆矩阵



附录 B RapidMiner

RapidMiner 是一种数据科学软件平台，可为数据准备、机器学习、深度学习、文本挖掘以及预测分析提供集成环境。它用于商务应用以及验证、教育、培训、快速原型和应用开发，并支持机器学习过程的所有阶段，包括数据准备、结果可视化、模型验证和优化。

本附录描述了使用 RapidMiner Studio 从 ARFF 文件生成决策树的过程。为此，考虑了洗手检测算法的简单示例。打开 RapidMiner Studio 之后，主要步骤如下：

1. 从左侧的 **[Operators]**（运算符）窗口添加 **[Open File]**（打开文件）运算符，并将该运算符拖到空白 **[Process]**（处理）窗口中，如图 53 中所示。
2. 双击 **[Open File]**（打开文件）运算符，以选择要载入的 ARFF 文件。
3. 找到 **[Read ARFF]**（读取 ARFF）运算符，并将其拖到 **[Process]**（处理）窗口中。然后将 **[Read ARFF]**（读取 ARFF）运算符连接到 **[Open File]**（打开文件）运算符，如图 54 中所示。
4. 找到 **[Set Role]**（设置角色）运算符并将其拖动到 **[Process]**（处理）窗口中。然后，双击 **[Set Role]**（设置角色）运算符，并在 **[Parameters]**（参数）窗口中键入属性名称和目标角色，如图 55 中所示。
5. 找到 **[Decision Tree]**（决策树）运算符并设置相应的参数，如图 56 中所示。还需要将 **[Decision Tree]**（决策树）运算符连接到 **[res]**。
6. 单击 RapidMiner Studio 左上方的 **[Run]**（运行）按钮（蓝色三角形图标）。
7. 单击 **[Run]**（运行）按钮后，**[Results]**（结果）选项卡将根据 **[Graph]**（图）（图 57）和 **[Description]**（说明）显示生成的决策树。
8. 在生成的决策树的 **[Description]**（说明）部分中（图 58），需要将决策树复制到文本文件中，以便能够导入到 Unico 中的 MLC 工具。

图 53. RapidMiner Studio——打开文件

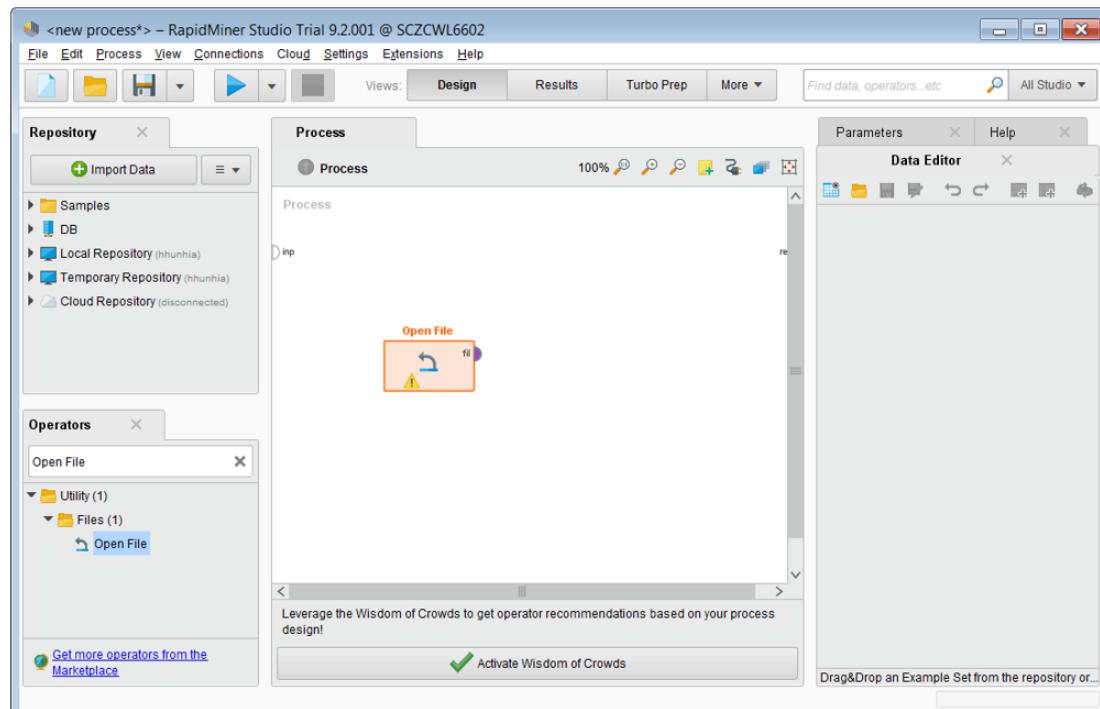


图 54. RapidMiner Studio——读取 ARFF

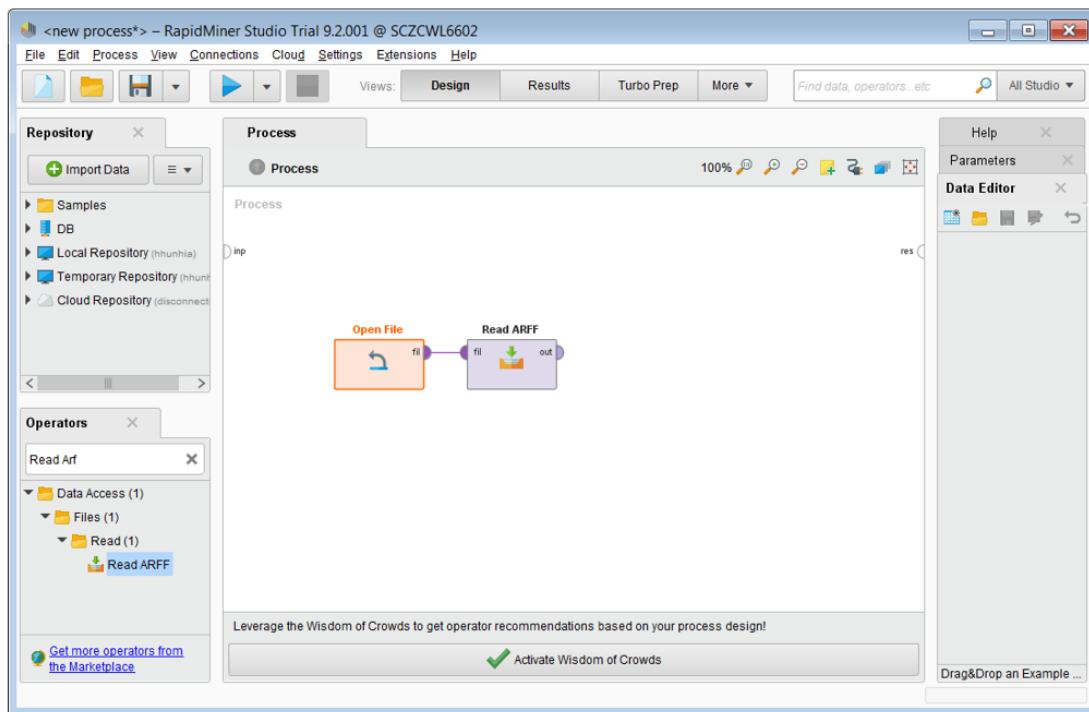


图 55. RapidMiner Studio——设置角色

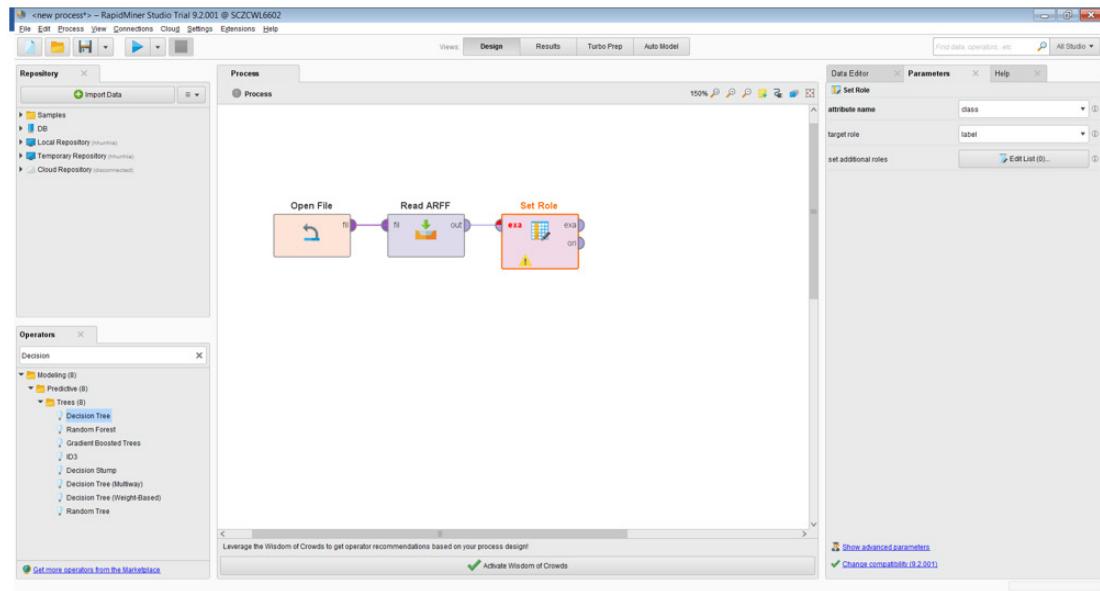


图 56. RapidMiner Studio——决策树运算符

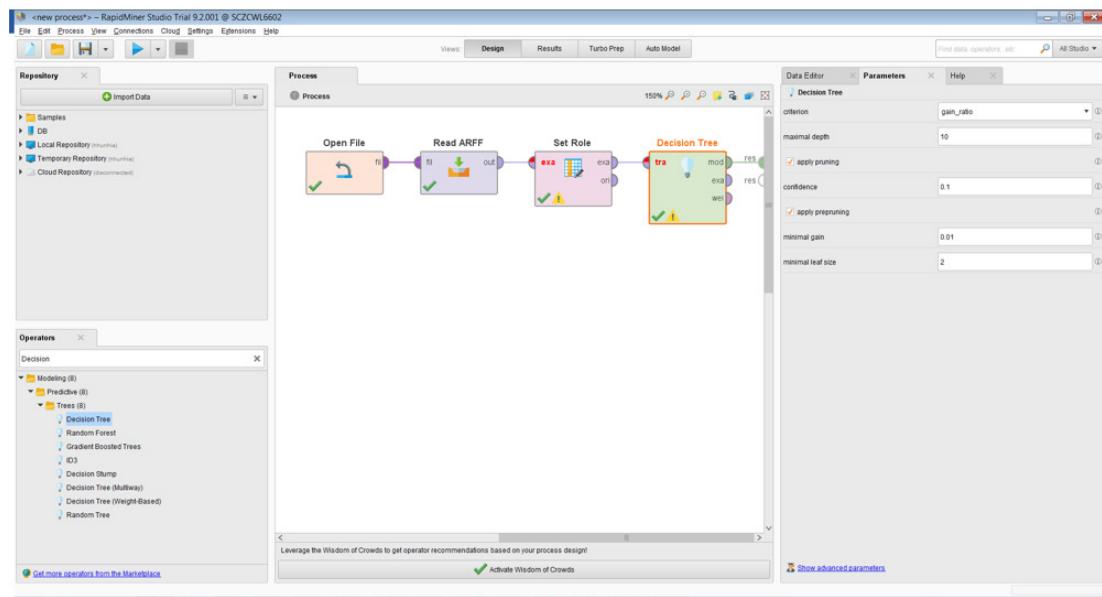


图 57. RapidMiner Studio——决策树图

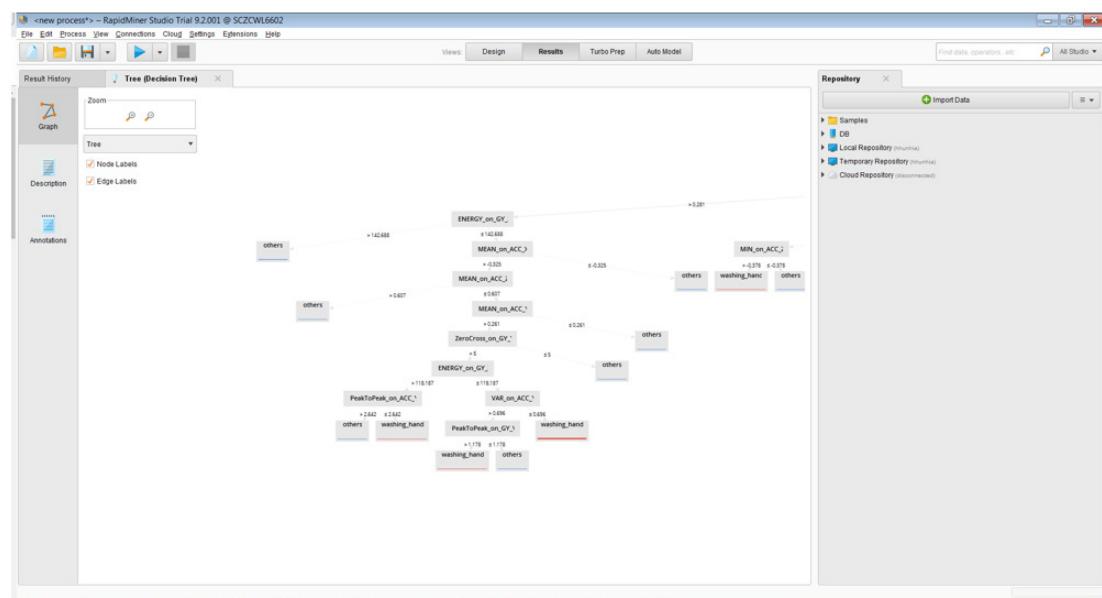
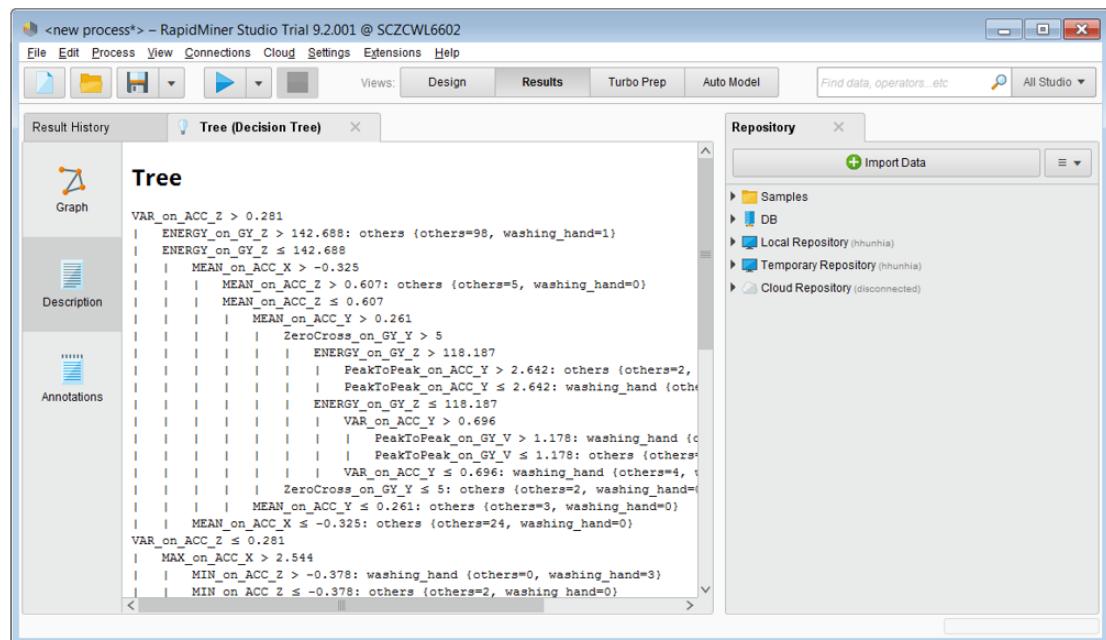


图 58. RapidMiner Studio——决策树说明



附录 C MATLAB®

可使用 MATLAB® 生成机器学习内核的决策树。MATLAB 上提供 MATLAB® 专用脚本。

将全部脚本导入 MATLAB® 工作区后，应调用 **[Generate_DcisionTree()]** 函数，并指定两个文件名（一个为 **.arff** 文件，其中包含通过 Unico 中的机器学习内核工具计算出的特征，另一个为 **.txt** 文件，其中包含生成的决策树）：

```
filename_ARFF = 'features.arff';
filename_dectree = 'decision_tree.txt';
Generate_DcisionTree(filename_ARFF, filename_dectree);
```

可在 GitHub 存储库的 **[matlab]** 文件夹所包含的 **README.md** 文件中找到更多详细信息。

附录 D Python

可通过“*scikit*”包使用 Python 生成机器学习内核的决策树。

Python 即可提供作为 Jupyter 记事本 (*.ipynb) 的 Python 脚本，也可以提供通用 Python 脚本 (*.py)。可在 GitHub 存储库的 [python] 文件夹所包含的 *README.md* 文件中找到更多详细信息。

附录 E 词汇表

本节包含用于机器学习的词汇表。大多数词汇表取自
[https://developers.google.com/machine-learning/glossary/。](https://developers.google.com/machine-learning/glossary/)

ARFF	ARFF（属性关系文件格式）文件是 ASCII 文本文件，描述了共享一组属性的实例列表。ARFF 文件由怀卡托大学计算机科学系的机器学习项目部开发，用于 Weka 机器学习软件。
J48	C4.5 算法的开源 Java 实现
Matplotlib	一种开源的 Python 2D 绘图库。matplotlib 可帮助您可视化机器学习的不同方面。
ML	机器学习的缩写。
ROC 曲线	ROC = 接收器的工作特征 真阳性率与假阳性率在不同分类阈值下的对比曲线。
Scikit-learn	流行的开源 ML 平台。参见 www.scikit-learn.org 。
Weka	一组用于数据挖掘任务的机器学习算法。它包含用于数据准备、分类、回归、聚类、关联规则挖掘和可视化的工具。
半监督式学习	根据数据训练模型，其中一些训练示例有标签，而其他示例无标签。一种半监督式学习技术是，为不带标签的示例推断标签，然后根据推断出的标签进行训练，以创建新模型。如果获取标签的成本很高，而不带标签的示例很多，则半监督式学习可能很有用。
标签	在监督式学习中，示例的“答案”或“结果”部分。带标签数据集中的每个示例均包含一个或多个特征及一个标签。例如，在房屋数据集中，特征可能包括卧室数量、浴室数量、房屋使用年限，而标签可能为房屋价格。在垃圾邮件检测数据集中，特征可能包括主题行、发件人和电子邮件本身，而标签可能为“垃圾邮件”或“非垃圾邮件”。
表示	将数据映射到有用特征的过程。
不带标签的示例	包含特征但不带标签的示例。不带标签的示例为要推断的输入。在半监督式和无监督式学习中，在训练过程中使用了不带标签的示例。
不平衡数据集	类别不平衡数据集的同义词。
参数	ML 系统自行训练的模型变量。
打分	推荐系统的一部分，为候选生成阶段所产生的每个项目提供一个值或排名。
代理标签	用于近似表示数据机中不直接可用的标签的数据。 例如，假设您将“会下雨吗？”作为数据集的布尔标签，但是该数据集不包含降雨数据。如果有照片，您可以将带伞的人的照片作为“会下雨吗？”的代理标签。但是，代理标签可能会使结果失真。例如，在某些地方，携带雨伞防晒比防雨淋更常见。
动态模型	以持续更新方式在线训练的模型。即，将数据连续输入模型。
独立且恒等分布 (i.i.d)	从恒定分布中提取的数据，每个绘制的值都不依赖于先前绘制的值。i.i.d 是一种理想的机器学习条件——是一种有用的数学构造，但在现实世界中几乎找不到。例如，网页的访客分布可能为 i.i.d。在短暂的时间内，即在那个短暂的窗口内分布不会改变，一个人的访问通常独立于另一个人的访问。但是，如果扩大了该时间窗口，则网页访问者的季节性差异可能会出现。
多类别分类	区分两个以上的类别的分类问题。例如，大约有 128 种枫树树种，因此对枫树树种进行分类的模型为多类别模型。相反，将电子邮件仅分为两类（垃圾邮件和非垃圾邮件）的模型为二进制分类模型。
多数类别	类别不平衡数据集中的更常见标签。例如，假定数据集包含 99% 的非垃圾邮件标签和 1% 的垃圾邮件标签，则非垃圾邮件标签为多数标签。
多项式分类	多类分类的同义词。
二进制分类	输出两个互斥类别之一的一种分类任务。例如，评估电子邮件消息，并输出“垃圾邮件”或“非垃圾邮件”的机器学习模型为二进制分类器。
泛化	是指模型对以前看不到的新数据（与用于训练模型的数据完全不同）做出正确预测的能力。
分类	标签的一组枚举目标值之一。例如，在检测垃圾邮件的二进制分类模型中，两种类别为垃圾邮件和非垃圾邮件。在识别狗的品种的多类别分类模型中，类别为贵宾犬、小猎犬、哈巴狗等。
分类模型	用于区分两个或多个离散类别的机器学习模型类型。例如，自然语言处理分类模型可确定输入的句子为法语、西班牙语还是意大利语。
分类阈值	一种标量值标准，该标准应用于模型预测分数，以将阳性类别和阴性类别分离。在将逻辑回归结果映射到二进制分类时使用。例如，考虑一个用于确定给定电子邮件消息是否为垃圾邮件的逻辑回归模型。如果分类阈值为 0.9，则逻辑回归值高于 0.9 的邮件被视为垃圾邮件，逻辑回归值低于 0.9 的邮件被视为非垃圾邮件。

服务	推断的同义词。
功能	用于进行预测的输入变量。
归一化	将实际值范围转换为标准值范围（通常为 -1 到 +1 或 0 到 1）的过程。例如，假设某个特征的自然范围为 800 到 6,000。通过减法和除法，您可以将这些值归一化为 -1 到 +1。另请参见缩放。
过拟合	创建的模型与训练数据的匹配度过高，以至于该模型无法对新数据做出正确预测。
回归模型	一种输出连续（通常为浮点）值的模型。相比之下，分类模型输出离散值。如“萱草”或“卷丹”。
混淆矩阵	NxN 表总结了分类模型预测的成功程度，即标签与模型分类的相关性。混淆矩阵的一个轴为预测模型标签，另一个轴为实际标签。
机器学习	通过输入数据构建（训练）预测模型的程序或系统。通过从训练模型时的相同分布获得的新（从未见过）数据，该系统使用学得的模型来做出有用的预测。机器学习还指与这些程序或系统有关的研究领域。
假阳性 (FP)	模型错误地预测阳性类别的示例。例如，该模型推断出特定电子邮件为垃圾邮件（阳性类别），但该电子邮件实际上不是垃圾邮件。
假阳性率 (FPR)	ROC 曲线中的 x 轴。假阳性率定义如下： $\text{假阳性率} = \text{假阳性} / (\text{假阳性} + \text{真阴性})$
假阴性 (FN)	模型错误地预测阴性类别的示例。例如，该模型推断出特定电子邮件不是垃圾邮件（阴性类别），但该电子邮件实际上为垃圾邮件。
监督式机器学习	根据输入数据及其相应的标签训练模型。监督式机器学习类似于学生通过研究一组问题及其相应的答案来学习一门学科。在掌握问题与答案之间的映射后，学生可以为同一主题的新问题（从未见过）提供答案。 与无监督式机器学习进行比较。
交叉验证	一种通过根据训练集中保留的一个或多个非重叠数据子集测试模型来估计模型泛化为新数据的程度的机制。
节点 (决策树)	对属性的“测试”。
精度	分类模型的指标。精度用于确定预测阳性类别时的模型正确频率。即： $\text{精度} = \text{真阳性} / (\text{真阳性} + \text{假阳性})$
静态模型	离线训练的模型。
决策边界	模型在二元类别或多类别分类问题中学到的类别之间的分隔符。
决策树	以分支语句序列表示的模型。
决策阈值	分类阈值的同义词。
可解释性	模型预测的解释难易程度。深度模型通常无法解释；也就是说，深层模型的不同层可能难以解读。相比之下，线性回归模型和广义模型通常更具可解释性。
类别不平衡数据集	两个类别的标签具有明显不同的频率的二进制分类问题。例如，0.0001 个示例具有阳性标记且 0.9999 个示例具有阴性标记的疾病数据集为类别不平衡问题，但 0.51 个示例标记一个球队获胜而 0.49 个示例标记另一个球队获胜的足球比赛预测器为类别不平衡问题。
离散特征	具有一组有限可能值的特征。例如，其值只能为动物、植物或矿物的特征为离散（或类属）特征。与连续特征形成对比。
模型训练	确定最佳模型的过程。
目标	标签的同义词。
判别模型	通过一组一个或多个特征预测标签的模型。更正式地讲，判别模型在给定特征和权重下定义了输出条件概率。
平稳性	数据集中的数据属性，其中数据分布在一个或多个维度上保持不变。最常见的是，如果维度是时间，则意味着呈现出平稳性的数据不会随时间而变化。例如，呈现出平稳性的数据从 9 月到 12 月没有变化。
评分员	在示例中提供标签的人。有时被称为“注解者”。
启发法	问题的快速解决方案，可能是最佳解决方案，也可能不是最佳解决方案。

欠拟合	由于模型未捕捉到训练数据的复杂性，所产生的模型的预测能力较差。许多问题可能会导致欠拟合，包括： <ul style="list-style-type: none">训练错误的特征集。训练时间太短或学习率太低。训练所采用的正规化率过高。在深度神经网络中提供的隐藏层太少。
强化学习	一种机器学习算法，该算法通过一系列操作后的反馈（奖励和惩罚）来实现最终奖励的最大化。例如，大多数游戏的最终奖励为胜利。通过评估最终导致获胜的先前游戏动作序列以及最终导致失败的序列，强化学习系统可成为玩游戏的专家。
少数类别	类别不平衡数据集中的较不常见标签。例如，假定数据集包含 99% 的非垃圾邮件标签和 1% 的垃圾邮件标签，则垃圾邮件标签为少数标签。
神经网络	一种从大脑获得灵感的模型，该模型分为多层（至少有一层是隐藏的），这些层包含简单的连接单元或神经元以及非线性元素。
实例	例子的同义词。
示例	一行数据集。包含一个或多个特征，也可能包含一个标签的示例。另请参见带标签示例和不带标签的示例。
属性/特征	属性为实例的一个方面（如温度、湿度）。在机器学习中，属性通常被称为特征。特殊属性为定义该示例所属类别的类别标签（在监督学习中是必须的）。
数据分析	通过考虑样本、测试与可视化来了解数据。在建立第一个模型之前首次接收到数据集时，数据分析可能非常有用。这对于了解实验和调试系统问题也至关重要。
数据扩增	通过转换现有示例来创建其他示例，可人为扩增培训示例的范围和数量。例如，假设图像是您的特征之一，但您的数据集没有供模型学习有用的关联的足够图像示例。理想情况下，您将为数据集添加足够的标签图像，以正确训练模型。如果这不可行，数据扩增可通过旋转、拉伸和反射每个图像来产生原始图片的多个变体，从而可能产生足够的带标签数据，以便实现出色的训练。
数据组或数据集	示例集。
数值数据	以整数或实数值表示的特征。
缩放	特征工程中的一种常用做法，即通过调整特征的值范围，以便与数据集中的其他特征的范围匹配。例如，假如您希望数据集中的所有浮点特征的范围为 0 到 1。给定的特定特征的范围为 0 到 500，您可以通过将每个值除以 500 来缩放该特征。 另请参见归一化。
特征工程	确定那一特征可能对训练模型有用并将日志文件和其他来源的原始数据转换为所述特征的过程。特征工程有时也称为特征提取。
特征集	机器学习模型训练的一组特征。例如，邮政编码、财产规模和财产状况可能包括一个用于预测房价的模型的简单特征集。
特征提取	具有以下定义之一的重载术语： <ul style="list-style-type: none">检索通过无监督或预先训练的模型计算出的中间特征表示，以用作另一个模型的输入。特征工程的同义词。
推论	在机器学习中，通常是指通过将训练后的模型应用于不带标签的示例来进行预测的过程。在统计学上，推论是指根据某些观察到的数据拟合分布参数的过程。
无监督式机器学习	训练模型以查找数据集中的模式，通常是不带标签的数据集。 无监督式机器学习的最常见用途为将数据聚合为类似示例组。例如，无监督式机器学习算法可根据音乐的各种属性将歌曲聚合在一起。生成的群集可成为其他机器学习算法（例如，音乐推荐服务）的输入。聚类在难以获得真实标签的领域可能会有所帮助。例如，在反滥用和反欺诈等领域，群集可以帮助人们更好地理解数据。 无监督式机器学习的另一个示例为主成分分析 (PCA)。例如，通过在包含数百万个购物车内容的数据集上应用 PCA，可能会发现，包含柠檬的购物车经常也包含抗酸剂。 与监督式机器学习进行比较。
下采样	可能表示以下任一情况的重载术语： <ul style="list-style-type: none">减少特征中的信息量，以更有效地训练模型。以不成比例的低百分比训练超代表性类别示例，以改善代表性不足的类别的模型训练。
限幅	一种用于处理异常值的技术。具体而言，将大于最大设置值的特征值减小为该最大值。同样，将小于特定最小值的特征值增加至该最小值。
线性回归	通过输入特征的线性组合输出连续值的一种回归模型。

性能	重载术语具有以下含义： <ul style="list-style-type: none">软件工程中的传统含义。即：该软件运行速度有多快（或效率有多高）？在 ML 中的含义性能在这里回答了以下问题：该模型有多正确？也就是说，模型的预测效果如何？
序列模型	其输入具有序列依赖性的模型。例如，根据之前观看的一系列视频来预测下一个观看的视频。
训练	确定组成模型的理想参数的过程。
训练集	用于训练模型的数据集的子集。 与验证集和测试集形成对比。
验证	作为训练的一部分，用于使用验证集评估机器学习模型质量的过程。由于验证集与训练集不相交，因此验证有助于确保模型的性能在训练集之外得以推广。 与测试集形成对比。
验证集	用于验证的数据集子集，该子集与训练集不相交。 与训练集和测试集形成对比。
阳性类别	在二进制分类中，将两种可能的类别分别标记为阳性和阴性。阳性结果是我们要测试的东西。（诚然，我们同时测量两种结果，但要配合）例如，医学测试中的阳性类别可能为“肿瘤”。电子邮件分类器中的阳性类别可能为“垃圾邮件”。 与阴性类别形成对比。
异常值	偏离大多数其他值的值。在机器学习中，以下任何值均为异常值： <ul style="list-style-type: none">具有高绝对值的权重。与实际值相差甚远的预测值。与平均值大约相差 3 个标准偏差以上的输入数据值。 异常值通常会在模型训练中引起问题。限幅为一种管理异常值的方式。
阴性类别	在二进制分类中，一种类别被称为阳性类别，另一种类型被称为阴性类别。阳性类别为我们寻找的类别，而阴性类别则为另一种可能。例如，医学测试中的阴性类别可能为“非肿瘤”。电子邮件分类器中的阴性类别可能为“非垃圾邮件”。另请参阅阳性类别。
预测	提供输入示例时的模型输出。
预训练模型	已经训练过的模型或模型组件。
噪声	广义上讲，任何会使数据集中的信号模糊的东西。噪声可通过多种方式引入数据中。例如： <ul style="list-style-type: none">评分员在标签上犯错误。人员和仪器错误记录或忽略了特征值。
召回率	分类模型的指标，可回答以下问题： “在所有可能的阳性标签中，模型可以正确识别多少个标签？” 即： $\text{召回率} = \text{真阳性} / (\text{真阳性} + \text{假阴性})$
真相	正确答案。现实。由于现实通常是主观的，因此评估专家通常为真相的代理。
真阳性 (TP)	模型正确地预测阳性类别的示例。例如，该模型推断出特定电子邮件为垃圾邮件，而该电子邮件确实为垃圾邮件。
真阳性率 (TPR)	召回率的同义词。即： $\text{真阳性率} = \text{真阳性} / (\text{真阳性} + \text{假阴性})$ 真阳性率是 ROC 曲线中的 y 轴。
真阴性 (TN)	模型正确地预测阴性类别的示例。例如，该模型推断出特定电子邮件不是垃圾邮件，而该电子邮件确实不是垃圾邮件。

版本历史

表 15. 文档版本历史

日期	版本	变更
2022 年 11 月 18 日	1	初始版本

目录

1 LSM6DSV16X 中的机器学习内核	2
1.1 输入	4
1.2 滤波器	7
1.2.1 滤波系数	8
1.3 特征	10
1.3.1 均值	11
1.3.2 方差	11
1.3.3 能量	11
1.3.4 峰峰值	11
1.3.5 过零信号	12
1.3.6 正过零信号	12
1.3.7 负过零信号	13
1.3.8 峰值探测器	13
1.3.9 正峰值探测器	14
1.3.10 负峰值探测器	14
1.3.11 最小值	15
1.3.12 最大值	15
1.3.13 递归特征	16
1.3.14 特征选择	18
1.4 决策树	22
1.4.1 LSM6DSV16X 中的决策树限制	23
1.5 元分类器	24
1.5.1 LSM6DSV16X 中的元分类器限制	24
1.6 有限状态机接口	25
1.7 FIFO 中的 MLC 数据	26
2 机器学习内核工具	27
2.1 Unico GUI	27
2.2 决策树生成	30
2.3 配置过程	32
3 决策树示例	39
3.1 运动强度	39
3.2 6D 位置识别	39
3.3 健身活动识别	41
3.4 示例汇总	42
附录 A Weka	43
附录 B RapidMiner	50
附录 C MATLAB®	54
附录 D Python	55
附录 E 词汇表	56

版本历史	60
表一览	63
图一览	64

表一览

表 1.	机器学习内核输出数据速率	2
表 2.	滤波系数	7
表 3.	滤波器系数示例	8
表 4.	特征	10
表 5.	第一组递归特征	16
表 6.	第二组递归特征	17
表 7.	决策树结果	22
表 8.	决策树中断	23
表 9.	LSM6DSV16X 中的决策树限制	23
表 10.	元分类器示例	24
表 11.	LSM6DSV16X 中的元分类器限制	24
表 12.	FIFO 中的数据	26
表 13.	健身活动识别配置	41
表 14.	示例汇总	42
表 15.	文档版本历史	60

图一覽

图 1.	机器学习内核监督式方法.....	1
图 2.	LSM6DSV16X 中的机器学习内核.....	2
图 3.	机器学习内核模块.....	3
图 4.	MLC 输入（加速度计）	4
图 5.	MLC 输入（陀螺仪）	4
图 6.	滤波器基本元素	7
图 7.	峰峰值	11
图 8.	过零信号	12
图 9.	正过零信号	12
图 10.	负过零信号	13
图 11.	峰值探测器	13
图 12.	正峰值探测器.....	14
图 13.	负峰值探测器.....	14
图 14.	最小值	15
图 15.	最大值	15
图 16.	单个特征的三个不同类别的分布	19
图 17.	显示两个特征和两个类别	20
图 18.	使用自动输出工具实现的排名	21
图 19.	决策树节点	22
图 20.	滤波器和特征的标识符	26
图 21.	Unico GUI.....	27
图 22.	机器学习内核工具——数据模式	28
图 23.	机器学习内核工具——配置	29
图 24.	Unico 中的决策树生成	30
图 25.	Weka Preprocess（预处理）	31
图 26.	Weka Classify（分类）	31
图 27.	决策树格式	32
图 28.	配置过程	33
图 29.	将结果分配给数据模式	33
图 30.	机器学习内核配置	34
图 31.	滤波器配置	34
图 32.	特征配置	34
图 33.	ARFF 生成	35
图 34.	ARFF 文件	35
图 35.	结果和决策树配置	36
图 36.	元分类器和器件配置	36
图 37.	创建配置文件	37
图 38.	Unico 加载配置	37
图 39.	Unico 数据窗口	37

图 40.	Unico——机器学习内核源寄存器	38
图 41.	运动强度决策树	39
图 42.	6D 决策树	40
图 43.	健身活动识别决策树	41
图 44.	ARFF 示例	43
图 45.	Weka GUI Chooser (Weka GUI 选择器)	44
图 46.	Weka Explorer (浏览器)	45
图 47.	Weka Explorer (浏览器) ——属性	46
图 48.	Weka 分类	47
图 49.	Weka 分类 J48	47
图 50.	Weka J48 分类器参数	48
图 51.	正确分类实例	49
图 52.	混淆矩阵	49
图 53.	RapidMiner Studio——打开文件	50
图 54.	RapidMiner Studio——读取 ARFF	51
图 55.	RapidMiner Studio——设置角色	51
图 56.	RapidMiner Studio——决策树运算符	52
图 57.	RapidMiner Studio——决策树图	52
图 58.	RapidMiner Studio——决策树说明	53

IMPORTANT NOTICE – READ CAREFULLY

STMicroelectronics NV and its subsidiaries ("ST") reserve the right to make changes, corrections, enhancements, modifications, and improvements to ST products and/or to this document at any time without notice. Purchasers should obtain the latest relevant information on ST products before placing orders. ST products are sold pursuant to ST's terms and conditions of sale in place at the time of order acknowledgment.

Purchasers are solely responsible for the choice, selection, and use of ST products and ST assumes no liability for application assistance or the design of purchasers' products.

No license, express or implied, to any intellectual property right is granted by ST herein.

Resale of ST products with provisions different from the information set forth herein shall void any warranty granted by ST for such product.

ST and the ST logo are trademarks of ST. For additional information about ST trademarks, refer to www.st.com/trademarks. All other product or service names are the property of their respective owners.

Information in this document supersedes and replaces information previously supplied in any prior versions of this document.

© 2022 STMicroelectronics – All rights reserved