

上海大学 2019 — 2020 学年 冬季学期

课程名称: 机器学习与力学

课 程 号: 01826249

授课教师: 胡国辉

论文名称: 护卫舰推进设备速度评估决策树

学 号: 17121632

姓 名: 翟晗锋

专 业: 理论与应用力学

成 绩: _____

日 期: 2020/2/10

评 语:

护卫舰推进设备速度评估决策树

翟晗锋

(上海大学力学与工程科学学院, 上海 200444)

摘要

护卫舰作为海军配备常规舰种之一, 需要强大的推进装置以供其动力需要。通常在评估护卫舰动力强度及灵活性中最重要的指标就是其速度。其中, 很多因素均可对护卫舰运行速度产生影响。本文分别对海军护卫舰发动引擎的杠杆位置、燃气轮机轴扭矩、螺旋桨扭矩、压缩机内外气压、气温等 17 个相关参数以及船体运行速度的数据进行处理收集, 并对船体速度进行三种不同方法分类: (i) 将高于数据集中平均速度定义为快速, 其余为慢速; (ii) 将速度划分为五个区间, 由此定义五个不同的等级; (iii) 直接将数据集中速度的种类导入运算, 分别对参数和不同分类下的速度数据导入 MATLAB 软件中利用决策树方法进行计算; 并通过决策树计算所得结果分析出护卫舰中不同参数对于船体运行速度的影响。由此得到结论, 并判断出推动装置中那些参数对于船体整体速度起到更大的作用。

关键词: 护卫舰; 燃气轮机; 决策树; 机器学习

1 引言

随着人类科技水平及交通工具的不断发展, 人类一直在追求制造更高效快速的交通工具来达到不同目的。同时, 在当今社会, 海洋资源也称为人们激烈竞争的资源之一, 海洋权需要有强大的海军来支持。而对于一个海军来说, 最为基础的设施就是护卫舰, 其可以支援强大的海军兵力。其中, 高速航行是护卫舰的必备“能力”, 或者是最为重要的指标之一。其中, 有很多的因素支撑了护卫舰灵活强大的动力系统。

作为一名力学学生, 分析可能的参数对于动力系统的影响是我们的关注点。由课堂上讲述的决策树的方法, 我们可以将不同的参数输入进行计算分析, 由此可以推断不同的因子对推进速度造成的影响。

2 背景概括

2.1 护卫舰

在现代海军中，护卫舰用于保护其他军舰和商船，尤其是作为两栖远征军，会给正在执行任务中的补给团体和商队的反潜战（ASW）战斗人员提供补给。同时，被称为“护卫舰”的船只有时也会归类于驱逐舰，巡洋舰，甚至是战舰。欧洲的一些海军，例如荷兰，法国，德国或西班牙的海军，其驱逐舰和护卫舰都使用“护卫舰”一词。



图1 英国购买的被称为“箭头140”的军舰

Fig.1. The warship being purchased by Britain is known as the Arrowhead 140.

2.2 推进系统

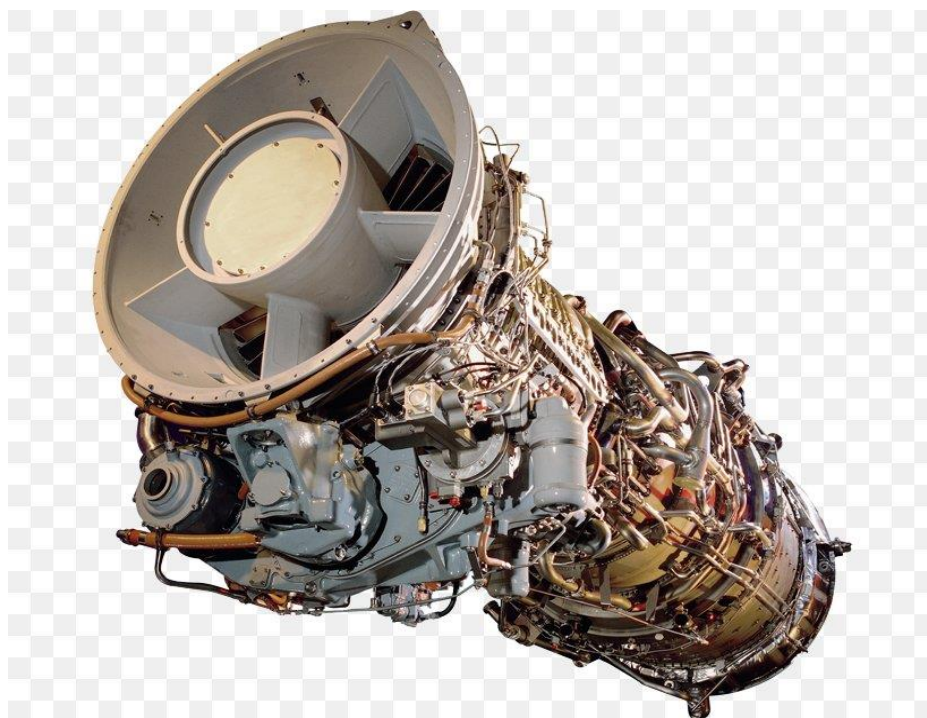


图 2 通用型电子 LM2500 燃气轮机柴油煤气船的燃气轮机

Fig.2. Ship - General Electric LM2500 Gas Turbine Combined Diesel And Gas Ship

燃气轮机推进系统起着重要作用，因为它推动了全球 80% 的活跃护卫舰舰队的发展，并且基本上所有大于 4,000T 的船舶都需要动力强劲的燃气轮机，其满足了护卫舰对速度和功率密度的需求。

同时，推进设备的尺寸和重量都是护卫舰关键属性。较小的设备可以给乘员，战斗系统和任务有效载荷提供更多空间。

3 方法

3.1 概念及定义

在研究分析数据集的纯度，我们经常选择信息熵做为度量指标。首先，我们先令集合 D 代表数据库，数据中第 k 类样本比例为 $p_k (k=1,2, \dots, |y|)$ ，那么现在信息熵 D 可定义为：

$$Ent(D) = - \sum_{k=1}^{|y|} p_k \log_2 p_k$$

其中， $Ent(D)$ 的数值越小， D 的纯度越高。

现假定离散属性 a 有 V 个可能的取值 $\{a^1, a^2, \dots, a^V\}$ ，现对于通用属性 a 对样本集 D 进行划分，那么会产生 V 个分支节点（第 v 个分支节点包含 D 所在属性 a 上取值为 a^v 的样本，我们记作 D^v ）。根据前述信息熵定义，可以考虑不同分支包含的不同样本数，同时赋予分支结点权重 $|D^v|/|D|$ ，则可解释或者推断出样本数越多的分支节点影响越大，即可计算属性 a 对于样本 D 的信息增益：

$$Gain(D, a) = Ent(D) - \sum_{v=1}^V \frac{|D^v|}{D} Ent(D^v)$$

在对样本集选择分类系数时，有很多种分类方法其中，本文中选择使用 Gini 系数进行分类与计算。首先，我们还可以用基尼值来分析或测量数据集 D 的纯度：

$$Gini(D) = \sum_{k=1}^{|y|} \sum_{k' \neq k} p_k p_{k'}$$

其中， $Gini(D)$ 的含义可理解为从数据集 D 中随机的抽取两个样本，他们的类别分别标记或代表了不一样的概率。所以，上述计算 $Gini(D)$ 值越小，数据集 D 纯度越高。其中，属性 a 基尼指数被定义为

$$Gini_index(D, a) = \sum_{v=1}^V \frac{|D^v|}{D} Gini(D^v)$$

3.2 问题分析

在该问题中，我们先需对数据集合变量进行分析。在数据集中，总共有 17 组自变量，这 17 组变量名称及单位如下表所示。

杠杆位置 []	燃气轮机轴扭矩 [kN·m]	引擎速率转数 [rpm]
气体发生器转数 [rpm]	右舷螺旋桨扭矩 [kN]	左舷螺旋桨扭矩 [kN]
高压 (HP) 涡轮出口温度 [C]	压缩机入口空气温度 [C]	燃气涡轮压缩机出口空气温度 [C]
HP 涡轮出口压力 [bar]	燃气涡轮压缩机入口空气压力 [bar]	燃气涡轮压缩机出口空气压力 [bar]
引擎排气压力 [bar]	涡轮喷射控制 [%]	燃料流量 [kg/s]
压缩机衰减状态系数	燃气涡轮衰减状态系数	船速 [knots]

表 1 数据集中的自变量和速度及其单位

Table 1. The decision tree results obtained from the second classification method.

在这 17 组变量中，我们需要分析出那些变量对于船速起到作用更为关键。因此，在进行决策树运算时，我们需要先对分析量速度进行分类：(i). 对于数据集中已有速度求平均值；大于平均值的定义为快速，数值为 1，小于平均值的定义为慢速，数值为 0. (ii). 将所有速度数值由小到大划分为五个等级，对五个等级分别赋数值 1, 2, 3, 4, 5. (iii). 将原有速度不改变，原始数值作为决策树的决定值。

对于第一种分类情况，我们需要分别计算 17 个自变量对应的基尼系数。在下列式中，*lever_position*为杠杆位置，*value₁*代表总数据集中具有杠杆位置 1(杠杆位置的第一个数值大小)的数据，而*value₁(high_speed)*代表在数值为“杠杆位置 1”大小的数据集中具有高速属性(前文中提到标记为数值 1 的属性)的数据；*value₂*代表杠杆位置 2 或杠杆位置这个数据集中第二个数值，同理，*value₂(h_s)*代表在该值中具有高速属性的数据；以此类推，*value_n*代表的是第 n 个数据。因我们有 17 个自变量或者属性，即统共可列 17 个方程。

$$\left\{ \begin{array}{l} Gini_index(D, lever_position) = \frac{value_1}{16400} \left\{ 2 \cdot \frac{value_1(high_speed)}{value_1} \cdot \left(1 - \frac{value_1(high_speed)}{value_1} \right) \right\} \\ + \frac{value_2}{16400} \left\{ 2 \cdot \frac{value_2(h_s)}{value_2} \cdot \left(1 - \frac{value_2(h_s)}{value_2} \right) \right\} + \dots + \frac{value_n}{16400} \left\{ 2 \cdot \frac{value_n(h_s)}{value_n} \cdot \left(1 - \frac{value_n(h_s)}{value_n} \right) \right\} \\ Gini_index(D, shaft_torque) = \frac{v_1(shaft\ torque\ value)}{16400} \left\{ 2 \cdot \frac{v_1(h_s)(s_t)}{v_1(s_t)} \cdot \left(1 - \frac{v_1(h_s)}{v_1} \right) \right\} + \dots \\ Gini_index(D, exit_temperature) = \frac{v_1(exit\ temperature\ value)}{16400} \left\{ 2 \cdot \frac{v_1(h_s)}{v_1} \cdot \left(1 - \frac{v_1(h_s)}{v_1} \right) \right\} + \dots \\ \dots (17\ equations \rightarrow due\ to\ 17\ coefficients\ or\ indices) \dots \\ Gini_index(D, fuel_flow) = \frac{v_1(fuel\ flow\ value)}{16400} \left\{ 2 \cdot \frac{v_1(h_s)}{v_1} \cdot \left(1 - \frac{v_1(h_s)}{v_1} \right) \right\} + \dots \end{array} \right.$$

在第二种分类情形中，我们将所有的速度分为五个区间并分别定义五个等级来进行判断。对于每一个等级，计算方法同第一种分类情况。不同的是，我们需要对五个等级都进行计算，最后得出的决策树可以判断对于每个等级速度起到关键作用的自变量。计算方法如下：

对于第一等级速度 (level 1 speed): 如下方程所示, *lever_position* 代表杠杆位置, *shaft_torque* 代表轴机扭矩, 其余还有 15 个自变量分别代表影响速度的不同因素；通过计算他们的 Gini 系数, 并可通过其数值大小判断出其对每个等级速度的影响力大小。

$$\left\{ \begin{array}{l} Gini_index(D, lever_position) = \frac{value_1}{16400} \left\{ 2 \cdot \frac{value_1(level_1)}{value_1} \cdot \left(1 - \frac{value_1(level_1)}{value_1} \right) \right\} \\ + \frac{value_2}{16400} \left\{ 2 \cdot \frac{value_2(l_1)}{value_2} \cdot \left(1 - \frac{value_2(l_1)}{value_2} \right) \right\} + \dots + \frac{value_n}{16400} \left\{ 2 \cdot \frac{value_n(l_1)}{value_n} \cdot \left(1 - \frac{value_n(l_1)}{value_n} \right) \right\} \\ Gini_index(D, shaft_torque) = \frac{v_1(shaft\ torque\ value)}{16400} \left\{ 2 \cdot \frac{v_1(l_1)(s_t)}{v_1(s_t)} \cdot \left(1 - \frac{v_1(l_1)}{v_1} \right) \right\} + \dots \\ \dots (17\ equations) \dots \\ Gini_index(D, fuel_flow) = \frac{v_1(fuel\ flow\ value)}{16400} \left\{ 2 \cdot \frac{v_1(l_1)}{v_1} \cdot \left(1 - \frac{v_1(l_1)}{v_1} \right) \right\} + \dots \end{array} \right.$$

对于第二等级速度 (level 2 speed):

$$\left\{ \begin{array}{l} Gini_index(D, lever_position) = \frac{value_1}{16400} \left\{ 2 \cdot \frac{value_1(level_2)}{value_1} \cdot \left(1 - \frac{value_1(level_2)}{value_1} \right) \right\} \\ + \frac{value_2}{16400} \left\{ 2 \cdot \frac{value_2(l_2)}{value_2} \cdot \left(1 - \frac{value_2(l_2)}{value_2} \right) \right\} + \dots + \frac{value_n}{16400} \left\{ 2 \cdot \frac{value_n(l_2)}{value_n} \cdot \left(1 - \frac{value_n(l_2)}{value_n} \right) \right\} \\ Gini_index(D, shaft_torque) = \frac{v_1(shaft\ torque\ value)}{16400} \left\{ 2 \cdot \frac{v_1(l_2)(s_t)}{v_1(s_t)} \cdot \left(1 - \frac{v_1(l_2)}{v_1} \right) \right\} + \dots \\ \dots (17\ equations) \dots \end{array} \right.$$

...

对于第五等级速度 (level 5 speed):

$$\left\{ \begin{array}{l} Gini_index(D, lever_position) = \frac{value_1}{16400} \left\{ 2 \cdot \frac{value_{1(level_5)}}{value_1} \cdot \left(1 - \frac{value_{1(level_5)}}{value_1} \right) \right\} \\ + \frac{value_2}{16400} \left\{ 2 \cdot \frac{value_{2(l_5)}}{value_2} \cdot \left(1 - \frac{value_{2(l_5)}}{value_2} \right) \right\} + \dots + \frac{value_n}{16400} \left\{ 2 \cdot \frac{value_{n(l_5)}}{value_n} \cdot \left(1 - \frac{value_{n(l_5)}}{value_n} \right) \right\} \\ \dots (17 \text{ equations}) \dots \\ Gini_index(D, fuel_flow) = \frac{v_1(fuel \ flow \ value)}{16400} \left\{ 2 \cdot \frac{v_{1(l_5)}}{v_1} \cdot \left(1 - \frac{v_{1(l_5)}}{v_1} \right) \right\} + \dots \end{array} \right.$$

比对五个等级速度中单独计算不同自变量的 Gini 系数大小，分析出对于每个等级速度影响最大的因素，即可以画出对于第二种分类方法中速度的决策树。

对于最后一种速度分类方法，该种速度分类方法中，我们不人为的去划分界限，而是使速度自然变化定义为分级方式，每一个量对应一个等级。而针对每一个等级，都可以通过计算 Gini 系数的方式判断出对其影响较大的因素或自变量。对于每一个速度数值，计算方法同分类方法 1。

对于速度 1(第一个速度取值):

$$\left\{ \begin{array}{l} Gini_index(D, lever_position) = \frac{value_1}{16400} \left\{ 2 \cdot \frac{value_{1(speed_1)}}{value_1} \cdot \left(1 - \frac{value_{1(speed_1)}}{value_1} \right) \right\} \\ + \frac{value_2}{16400} \left\{ 2 \cdot \frac{value_{2(s_1)}}{value_2} \cdot \left(1 - \frac{value_{2(s_1)}}{value_2} \right) \right\} + \dots + \frac{value_n}{16400} \left\{ 2 \cdot \frac{value_{n(s_1)}}{value_n} \cdot \left(1 - \frac{value_{n(s_1)}}{value_n} \right) \right\} \\ \dots (17 \text{ equations}) \dots \end{array} \right.$$

...

对于速度 n(第 n 个速度取值):

$$\left\{ \begin{array}{l} Gini_index(D, lever_position) = \frac{value_1}{16400} \left\{ 2 \cdot \frac{value_{1(speed_n)}}{value_1} \cdot \left(1 - \frac{value_{1(speed_n)}}{value_1} \right) \right\} \\ + \frac{value_2}{16400} \left\{ 2 \cdot \frac{value_{2(s_n)}}{value_2} \cdot \left(1 - \frac{value_{2(s_n)}}{value_2} \right) \right\} + \dots + \frac{value_n}{16400} \left\{ 2 \cdot \frac{value_{n(s_n)}}{value_n} \cdot \left(1 - \frac{value_{n(s_n)}}{value_n} \right) \right\} \\ \dots (17 \text{ equations}) \dots \end{array} \right.$$

同上文，分别计算每个速度取值的 Gini 系数，我们可以获得该种分类方法的决策树。

3.3 分类计算

对于第一种方法，迭代计算后的结果如下；解得第一个自变量，即是杠杆位置 lever position 为速度的最大影响因素；即我们可以说这个自变量决定了速度到底定义为快速(high speed; 本文中定义为 1)还是慢速(low speed; 0).

Decision tree for classification

```
1  if x1<1.613 then node 2 elseif x1>=1.613 then node 3 else 1
2  class = 0
3  class = 1
```

表 2 第一种速度分类方法决策树计算结果

Table 2. The decision tree results obtained from the first classification method

同时我们还好奇，除了杠杆位置，哪个因素对于“快速”和“慢速”的划分也起到重要的位置；因此，我们去除掉“杠杆位置”这个自变量，利用剩下的 16 组自变量再次进行迭代计算，计算结果如下；根据结果可见，对于划分速度为快速慢速的决定因素第二重要的是第 4 个自变量，即气体发生器的转速。

Decision tree for classification

```
1  if x4<16.9605 then node 2 elseif x4>=16.9605 then node 3 else 1
2  class = 0
3  class = 1
```

表 3 第一种速度分类方法决策树计算结果

Table 3. The decision tree results obtained from the first classification method

同理，我们对第二种分类方法进行入上述过程计算，计算结果如下表。计算结果较为繁杂，我们仅能判断出部分信息，例如第 7 个自变量(高度涡轮出口温度)对于第 5 等级的划分起到更重要的比重；且每个自变量中存在前提和迭代关系；具体分析需要我们描绘出决策树才可更直观的分析结果，见下节。

Decision tree for classification

```
1  if x7<805.132 then node 2 elseif x7>=805.132 then node 3 else 5
2  if x1<3.6525 then node 4 elseif x1>=3.6525 then node 5 else 4
3  class = 5
4  if x1<2.616 then node 6 elseif x1>=2.616 then node 7 else 2
5  class = 4
6  if x1<1.613 then node 8 elseif x1>=1.613 then node 9 else 2
7  class = 3
8  class = 1
9  class = 2
```

表 4 对于第二种速度分类形式的决策树结果

Table 4. The decision tree results obtained from the second classification method.

同理，第三种分类方法结果见下表，我们依旧很难通过数据判断出那些变量分别对速度数值起到什么样的影响，但是我们可以大体判断速度共有 27 个不同数值， x_1 起到较大作用。具体分析见下节。

Decision tree for classification

- 1 if $x_1 < 7.677$ then node 2 elseif $x_1 \geq 7.677$ then node 3 else 6
 - 2 if $x_1 < 6.6615$ then node 4 elseif $x_1 \geq 6.6615$ then node 5 else 6
 - 3 if $x_1 < 8.753$ then node 6 elseif $x_1 \geq 8.753$ then node 7 else 24
 - 4 if $x_1 < 5.6575$ then node 8 elseif $x_1 \geq 5.6575$ then node 9 else 6
 - 5 class = 21
 - 6 class = 24
 - 7 class = 27
 - 8 if $x_1 < 2.616$ then node 10 elseif $x_1 \geq 2.616$ then node 11 else 6
 - 9 class = 18
 - 10 if $x_1 < 1.613$ then node 12 elseif $x_1 \geq 1.613$ then node 13 else 6
 - 11 if $x_1 < 3.6525$ then node 14 elseif $x_1 \geq 3.6525$ then node 15 else 9
 - 12 class = 3
 - 13 class = 6
 - 14 class = 9
 - 15 if $x_1 < 4.6505$ then node 16 elseif $x_1 \geq 4.6505$ then node 17 else 15
 - 16 class = 12
 - 17 class = 15
-

表 5 对于第二种速度分类形式的决策树结果

Table 5. The decision tree results obtained from the second classification method.

4 结果

4.1 决策树

根据上节所述，我们已经通过计算 Gini 系数大小判断了对于每种分类方法的那些自变量对于影响或者判断速度大小起到了决定作用；接下来，我们需要将上节中的计算结果可视化，绘制成决策树，在通过决策树分析判断这些自变量对于船体运行速度的影响。

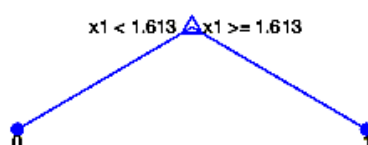


图3 第一种分类方法的决策树

Fig.3. Decision tree for the first classification method.

根据上图，我们也可以很明显的得出上节相同的结论，即第一个自变量，杠杆位置，对于判定速度为快速或者慢速起到更为重要的因素。

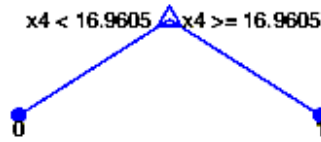


图4 第一种分类方法的决策树

Fig.4. Decision tree for the first classification method.

同时，如上图所示，去掉第一个自变量后，利用剩余自变量计算所得结果画成的决策树中，我们可以更加清晰的看出，我们如何通过第 4 个自变量，气体发生器转速，不同的数值来判定快速与慢速的定义界限。

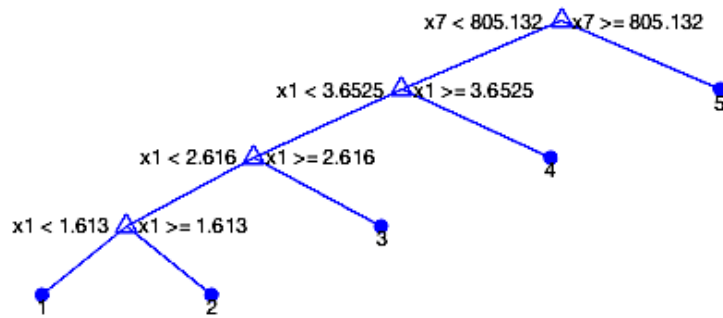


图5 第二种分类方法的决策树

Fig.5. Decision tree for the second classification method.

如上图，第二种速度分类方法的决策树非常直观为我们展示第 1 个自变量（杠杆位置）和第 7 个自变量（涡轮出口温度）如何通过不同数值范围对 5 个速度等级定义。有决策树可以判断出，当涡轮出口温度大于 805.132[C]，我们即可定义速度为第五等级；在涡轮出口温度小于该值时，当杠杆位置大于等于 3.6525，我们即定义此时速度为第 4 等级；以此类推，由决策树，我们可以判断出每个等级速度定义的前提条件。

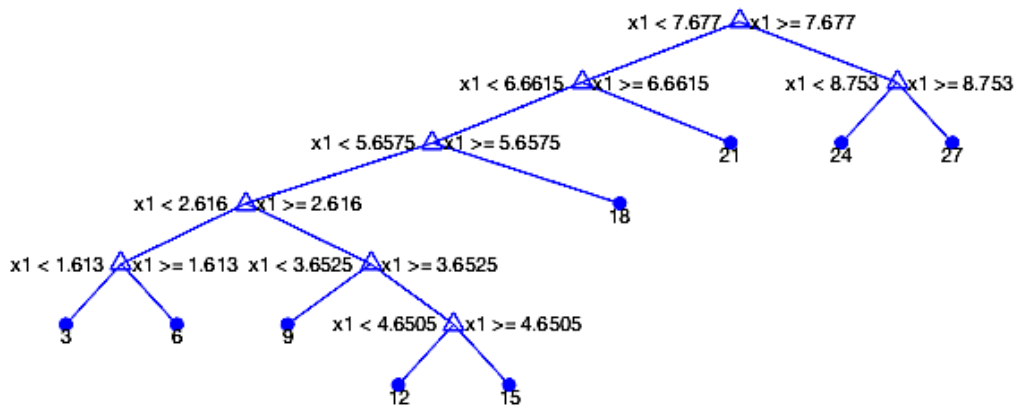


图6 第三种分类方法的决策树

Fig.6. Decision tree for the third classification method.

如上图，同理，对于第三种较为复杂的分类方法，我们将决策树绘制后，数据关系清晰可视。对于 27 个速度数值，影响最大的竟都是杠杆位置，在杠杆位置不同时，不同的数值条件下，会使得速度分别转向 27 个数值。

4.2 数据可视化

同时，为了更好的分析数据中大量的自变量之间的关系并使它们清晰显现，我利用数据可视化的方式，通过图片展现的方式大体判断出每个自变量对于最终战舰运行速度的影响；同时还可以判断自变量之间的关系。

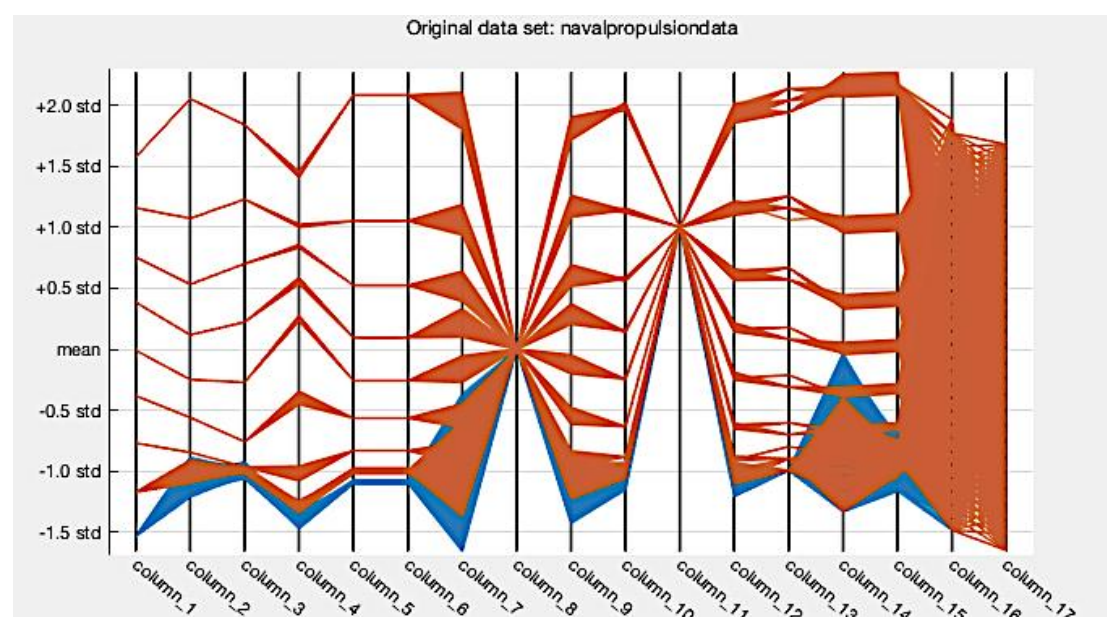


图7 第一种分类方法的数据可视化；其中，橙黄色为高速(1)，蓝色为低(0)
Fig.7. Data visualization for the third classification method. Among the statistics, the burnt orange area indicates high speed (1), the blue area indicates the low speed (0).

如图所示，我们可以判断出大量分布速度是大于平均速度，且我们可以看出前 6 个自变量基本是成正比关系的；更为重要的是，在区分快速和慢速这个界定界限上面，第 1、4、5、6、10 组自变量的数据均有较好的区分效果。这也说明判定军舰运行速度是否大于平均速度的指标其实有很多。

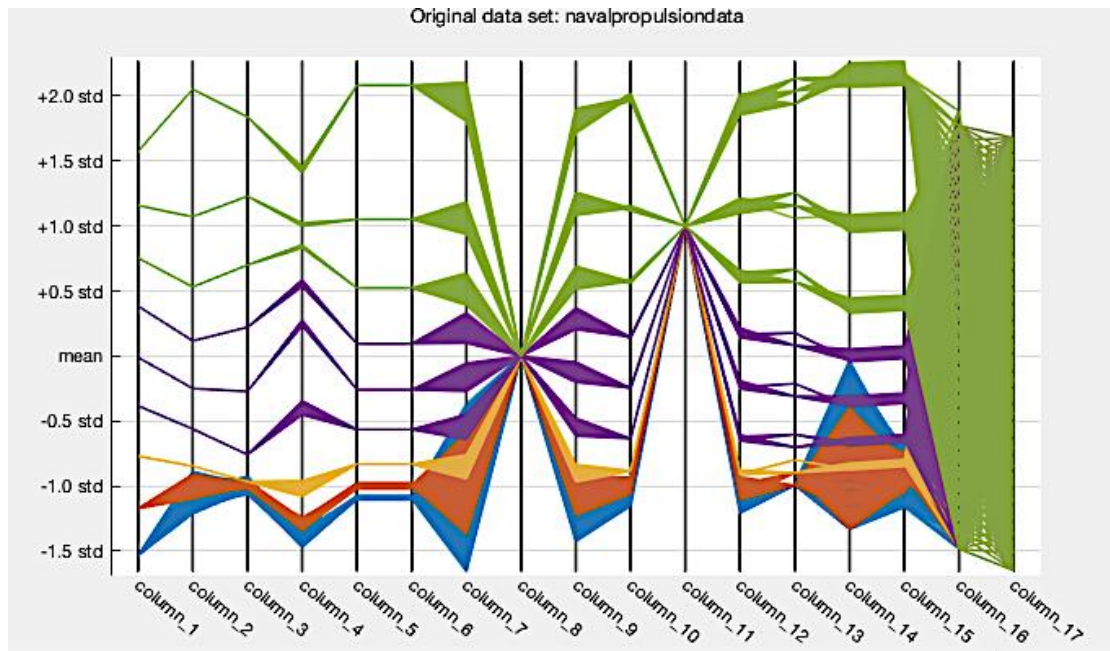


图8 第二种分类方法的数据可视化; 其中, 蓝色为等级1(1), 橙色为等级2(2), 黄色为等级3(3), 紫色为等级4(4), 绿色为等级5(5)。

Fig.8. Data visualization for the second classification method. Among the statistics, blue indicates level 1 (1), orange indicates level 2 (2), yellow indicates level 3 (3), purple indicates level 4 (4), green indicates level 5 (5).

由第二种速度分类的数据可视图, 得到的结论于第一种相似; 我们可以通过数据排列观察得第 1、4、5、6、7、9、10、12 均对区分不同等级速度展示出很好的效果, 这些自变量也可用于参考和判定。

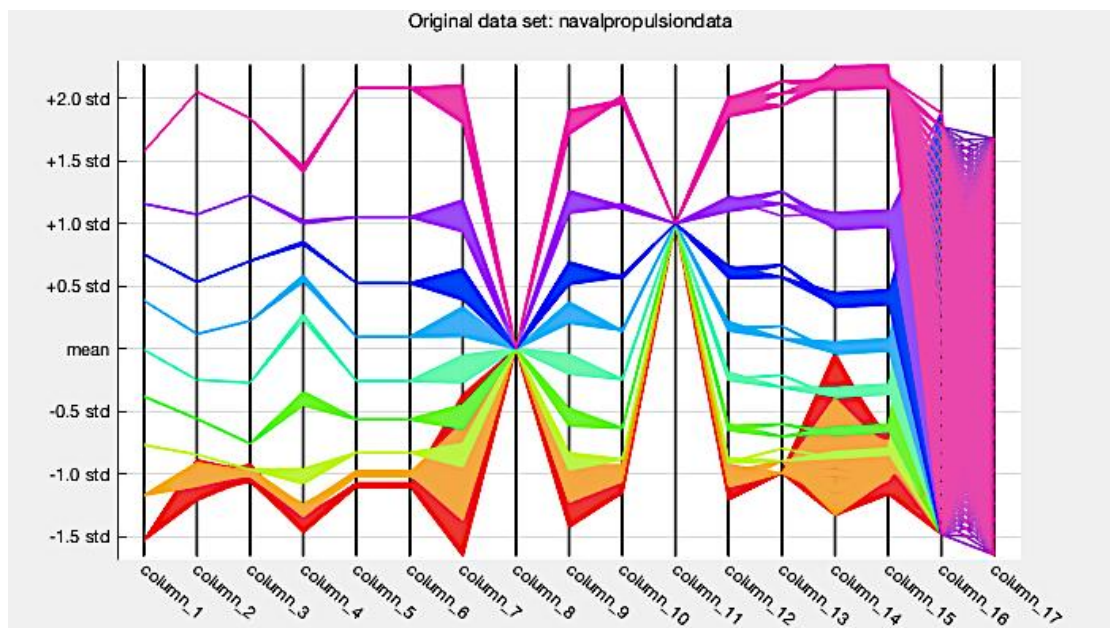


图9 第三种分类方法的数据可视化。每种颜色代表一种速度的取值

Fig.9. Data visualization for the third classification method. Each color indicates a speed value.

同上述两种分类方法，第三种对于速度更加精确的分类可视数据图让我们更加清晰的观测到每一种自变量或因子对于区分精确速度数值的权重或效果。结论大体同上：第 1、4、5、6、9、10 均可较为清晰的区分不同速度值。同时，我们也可由图判断前 6 组数据间呈现较好正比关系；此外，9、10 和 12、13 也呈现较为清晰的正比关系，可知在引擎发动时这些量直接可能因为机械的联动有着相应的关系。

5 结论

在护卫舰的推动系统中，很多因素都可对船体运行速度产生明显影响。通过对速度进行三种不同的划分(分类)方式，利用决策树的方法，我们可以判定推进系统中不同的因素对于速度的决定作用。(i) 当利用平均值将速度划分为两个区间时，当杠杆位置大于等于 1.613 时，速度为快速(定义为 1)，当小于该值，为慢速；同时，去掉杠杆位置这个因素后，决策树显示，当气体发生器转速大于等于 16.9605 转每分钟时，为快速，当小于该值，为慢速。(ii) 对于第二种分类方法，将速度由小到大划分为五个区间时，当高压涡轮出口温度大于等于 805.132 时，为第 5 等级速度；当起小于该值时，同时当杠杆位置大于等于 3.6525 时，为第四等级速度；当杠杆位置小于 3.6525 同时大于等于 2.616 时，为第三等级速度；当杠杆位置小于 2.616 同时大于等于 1.613 时，为第二等级速度；当起小于 1.613 时，为第一等级速度。(iii) 当速度的每一个取值为一个单独分类时，当杠杆位置大于等于 7.677 前提下，若其大于等于 8.753，此时为第 27 个速度取值；若其小于 8.753，为第 24 个速度取值。当其在小于 7.677 前提下，若大于等于 6.6615，为第 21 个速度取值，若小于 6.6615，在接下来仍有剩余小前提再次细分使得在不同的杠杆位置数值区间会得到不同的船体运行速度大小。

6 总结

本文引用了 UCI 机器学习数据库中由意大利 Università degli Studi di Genova 提供的海军推进设备的状态数据作为数据库，作者同时对数据重新做了分类处理、格式转换等其他方面的处理，但并未改变原有数据的真实性，以便

于在 MATLAB 平台上进行决策树的运算。基于决策树，在总共 17 组参数种筛选出对于速度影响最大的变化因子，并且根据不同的速度分类方式可以基于不同自变量画出不同的决策树。同时，该组数据尽管有 17 组自变量，也有可能并不充足，在实际舰船运行中，可能还有更多的因素影响。但本文提供了一种分析舰船性能，甚至研发更好更快舰船的可行性思路。

感谢上海大学力学与工程科学学院胡国辉教授的讲授与讨论。

参考文献

- [1] GE's Marine Solutions One Neumann Way MD S156 Cincinnati, Ohio USA 45215 www.ge.com/marine GE Marine Gas Turbines for Frigates. March 2018
- [2] A. Coraddu, L. Oneto, A. Ghio, S. Savio, D. Anguita, M. Figari, Machine Learning Approaches for Improving Condition-Based Maintenance of Naval Propulsion Plants, Journal of Engineering for the Maritime Environment, 2014, DOI: 10.1177/1475090214540874, (In Press)
- [3] P. Harrington(compile), Li.Y, Li.P, Qu.Y, Wang.B(translate)Machines Learning in Action 2013 (The People's Posts and Telecommunications Press) (in Chinese) [P. Harrington 编，李锐等 译 机器学习实战 2013 (人民邮电出版社)]
- [4] Zhou Zhihua. Machines Learning 2018 (Tsinghua University Press) (in Chinese) [周志华 机器学习 2018 (清华大学出版社)]
- [5] Joshi A.V. (2020) Decision Trees. In: Machine Learning and Artificial Intelligence. Springer, Cham
- [6] Mehmed Kantardzic. DECISION TREES AND DECISION RULES. Published Online: 25 October 2019 Published Print: 17 October 2019. Online ISBN: 9781119516057; Print ISBN: 9781119516040 DOI: 10.1002/9781119516057.ch6

Decision Tree Estimation on Frigate Propulsion Velocity

Zhai Hanfeng

(School of Mechanics and Engineering Science, Shanghai University, Shanghai 200444, China)

Abstract

As one of the most widely applied warships, frigate requires a high power propulsion system to support its heavyweight and nimble actions. Speed is one of the most important standards for frigates. What's more, there are many factors and variables that contribute to the highly agile movement of a frigate. Decision tree is a tool or a method to help people to identify which strategy most likely to reach a goal. This tool is a way to display an algorithm that only contains conditional control statements and it's very popular among machine learning. In this article, I use different sets of data obtained from frigate such as lever position, turbine shaft torque, rate of revolutions, speed, etc. Speed is the determining factor in decision making. To simplify the analysis, I use three methods to distinguish different speed: (i) Define the speed above average value as "high speed", and the rest are considered as low speed. (ii) set five sections within the whole speed value range and name them as five levels. (iii) input the statistics and define each speed value as a classification. By implementing these data into MATLAB And build a decision tree by calculating the Gini indices of each factor, it can be deduced that which factor plays a more important role while determining the speed of a frigate. Therefore, by applying these conclusions and knowledge, we might give rise to more ideas and inspiration in engineering and manufacturing.

Keywords: Frigate, Gas Turbine, Decision Tree, Machine Learning