

# 宠物狗推荐专家系统

## 一、选题背景

狗是人类的好伙伴。最早人们饲养狗主要是出于狩猎、放牧、看家等功能性目的，所以大多是体型较大的功能犬。随着时代的进步和发展，更多人开始出于单纯喜欢、作为陪伴，或者赶潮流而养狗。

但是我们身边经常出现由于没有经验，购买时仅根据外表漂亮而挑选，不了解其性格，或者不适合自己的日常生活习惯（例如有洁癖但是养了味道大爱掉毛的狗；周围有心脏病患者却养了爱叫的狗；上班族或者学生党养了精力过于旺盛需要运动的狗；以及房屋空间小却养了体型较大的狗）而后悔的现象。

为了解决这种问题，设计了本专家系统，充当一个在宠物狗选购领域较为专业或者有相关经验的专家，通过询问用户对宠物狗哪些属性参数更为在意，为用户推荐最为合适的宠物狗种类，尽量做到对狗狗负责，让用户不后悔。

## 二、数据采集

专家系统的构建需要有知识库作为支撑，本专家系统的数据采集主要来自：

- 窝狗网 (<http://www.wgpet.com/>)
- 狗扑网 ([www.goupu.com.cn/](http://www.goupu.com.cn/))

通过对身边同学的口头询问和一定数量的问卷调查，我们分析了大家最为关心的问题，再结合网络上可以获得的数据，最后决定采集以下属性作为衡量指标：

- 身高、体重（需要取平均值）
- 掉毛程度、体味程度、口水程度、美容程度、运动时间、关爱需求度
- 对小孩友善度、对生人友善度、对动物友善度
- 可训练度
- 忠诚度以及护卫能力
- 初养适应度
- 犬叫程度
- 颜值
- 搞笑程度
- 价格

最后向用户询问时的指标有 **10 个**：（最后打分时每个指标占 10 分）

- 体型（身高+体重）
- 打理照顾时间（掉毛程度+体味程度+口水程度+美容程度+运动时间+关爱需求度）
- 友善度（对小孩友善度、对生人友善度、对动物友善度）
- 可训练度
- 忠诚度以及护卫能力
- 初养适应度
- 犬叫程度
- 颜值
- 搞笑程度
- 价格

### 三、模糊专家系统设计

#### ● 设计思路

使用模糊专家系统对每一种狗计算出一个推荐度（打分），  
然后按照用户的排序来给每个指标推荐度乘上一个权重，获得最后的推荐的种类。  
具体需要使用模糊逻辑来计算的是前三个：体型、时间和友善度  
其余七个因为获取的数据中直接包含，所以不需要经过模糊逻辑的计算，计算时直接代入

#### ● 系统的实现

借助 matlab 中内置的模糊逻辑工具 Fuzzy Logic Toolbox 来实现模糊专家系统，默认使用的是 Mamdani 方法



图 1 模糊推理过程图

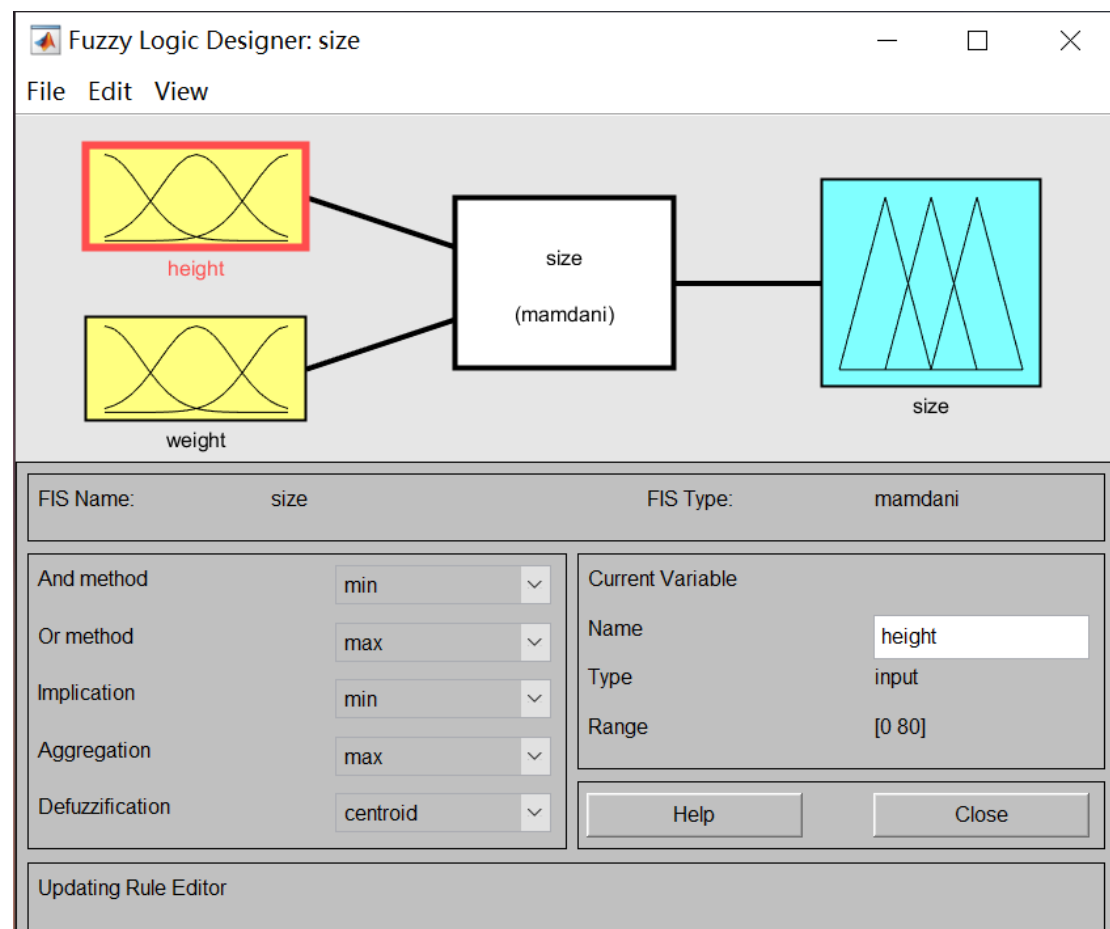
以体型大小判断为例说明

#### 步骤 1：模糊化

目的：取得清晰的输入，确定每个输入属于每个合适模糊集的程度

在 matlab 下输入 fuzzy 命令，打开 FIS 编辑器，设置系统的输入输出参数

体型判断的输入变量为身高和体重



打开隶属度函数编辑器为变量选择隶属度函数类型，给函数命名，并设置好参数

身高

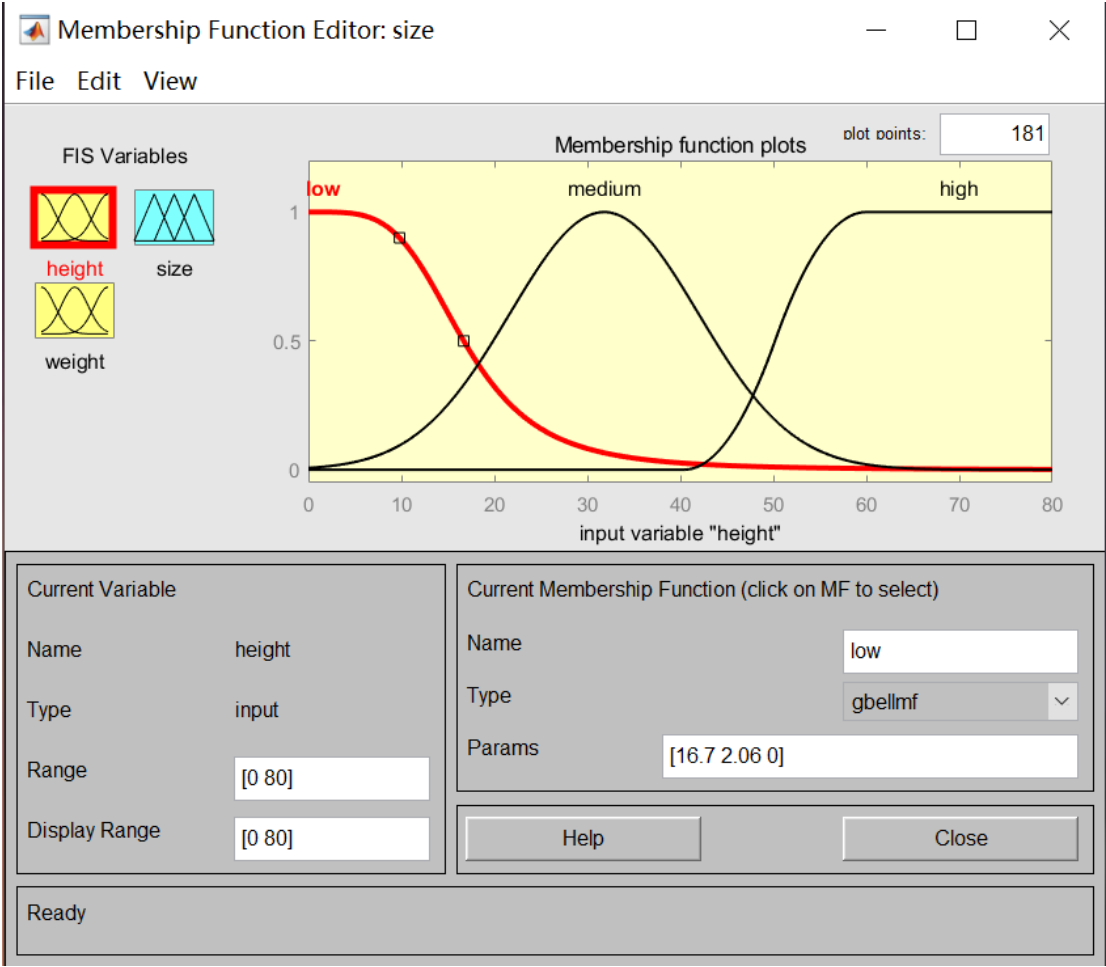
定义语言变量

参数范围根据收集到的数据得到最大值

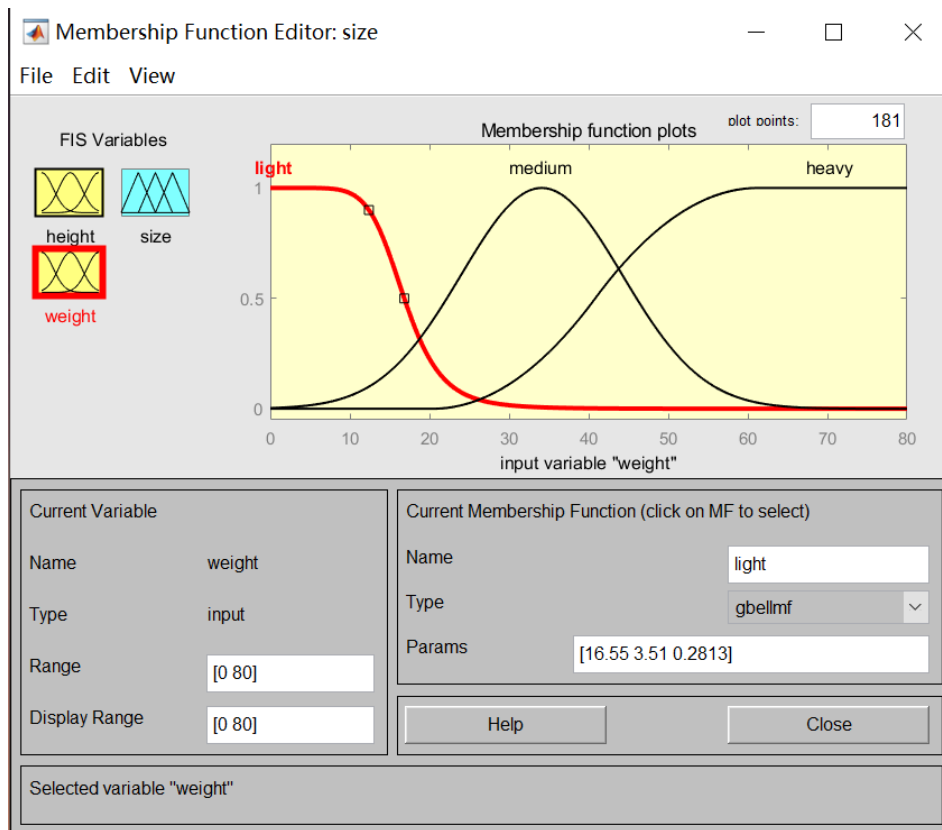
语言变量：身高（cm）	
语言值	值范围
low	[0,42]
medium	[0,70]
high	[40,80]

确定隶属度函数

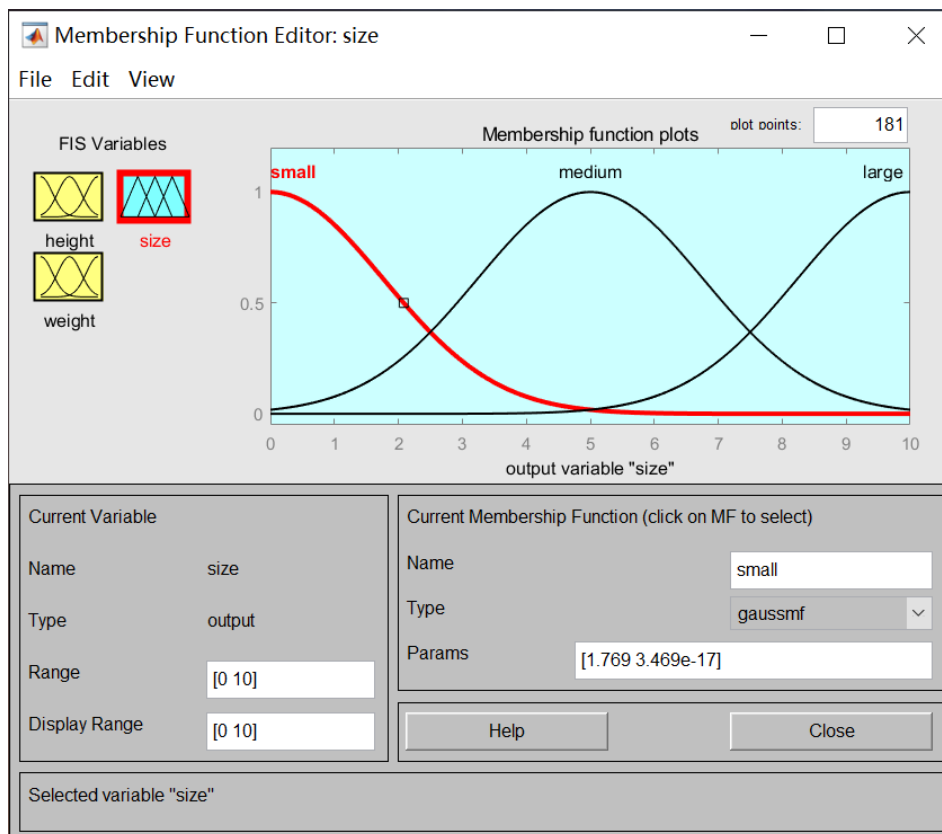
函数形状和位置参数由专家根据经验确定



## 体重



## 体型



步骤 2：规则评估

打开规则编辑器，可以看到模糊输入变量和输出变量的隶属度函数表，以及连接模糊输入变量间的模糊算子 and 和 or。根据规则表中的每一条规则选择前件和后件，添加即可  
这里的规则由我作为专家根据经验来制定，例如：

- 如果身高高，则体型大
- 如果身高矮，则体型小
- 如果身高中等，并且体重重，则体型大
- 如果身高中等，并且体重中等，则体型中等
- 如果身高中等，并且体重轻，则体型中等

Rule Editor: size

File Edit View Options

1. If (height is high) then (size is large) (1)

2. If (height is low) then (size is small) (1)

3. If (height is medium) and (weight is heavy) then (size is large) (1)

4. If (height is medium) and (weight is medium) then (size is medium) (1)

5. If (height is medium) and (weight is light) then (size is medium) (1)

If

height is

low  
medium  
high  
none

☐ not

and

weight is

light  
medium  
heavy  
none

☐ not

Then

size is

small  
medium  
large  
none

☐ not

Connection

☐ or

☒ and

Weight:

1

Delete rule

Add rule

Change rule

<<

>>

The rule is deleted

Help

Close

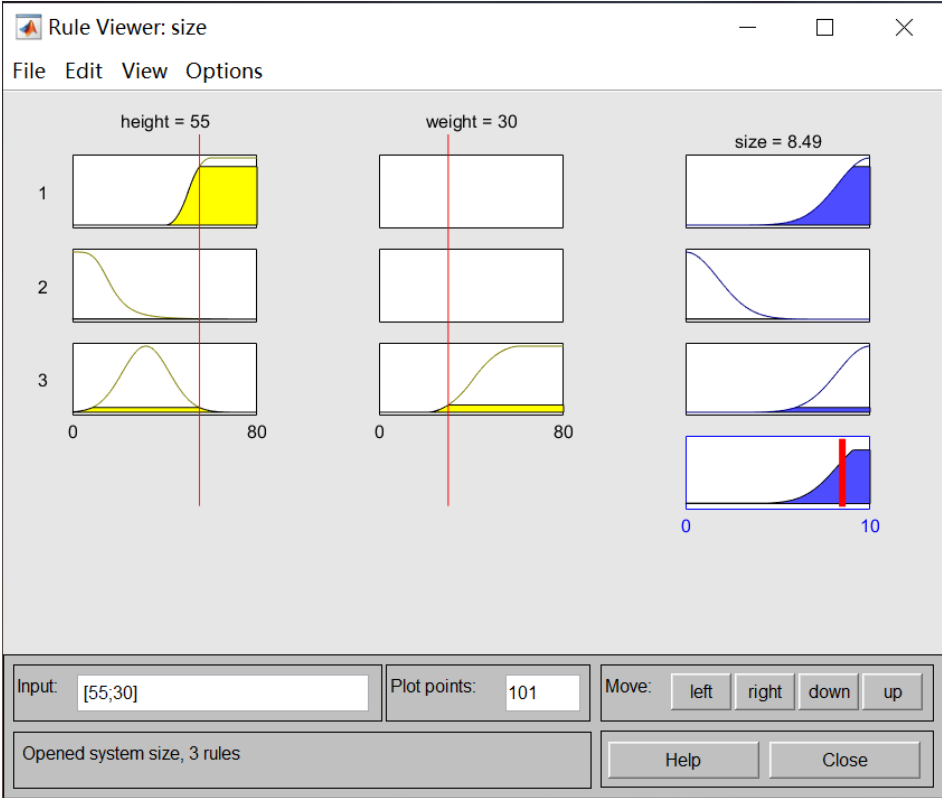
步骤 3：逆模糊化

模糊化可以帮助我们评估规则，但是模糊系统的最终输出必须是一个清晰的值。逆模糊化过程输入是集成的模糊输出集，输出是单一的数值，采用的是质心技术(centroid technique)

Defuzzification

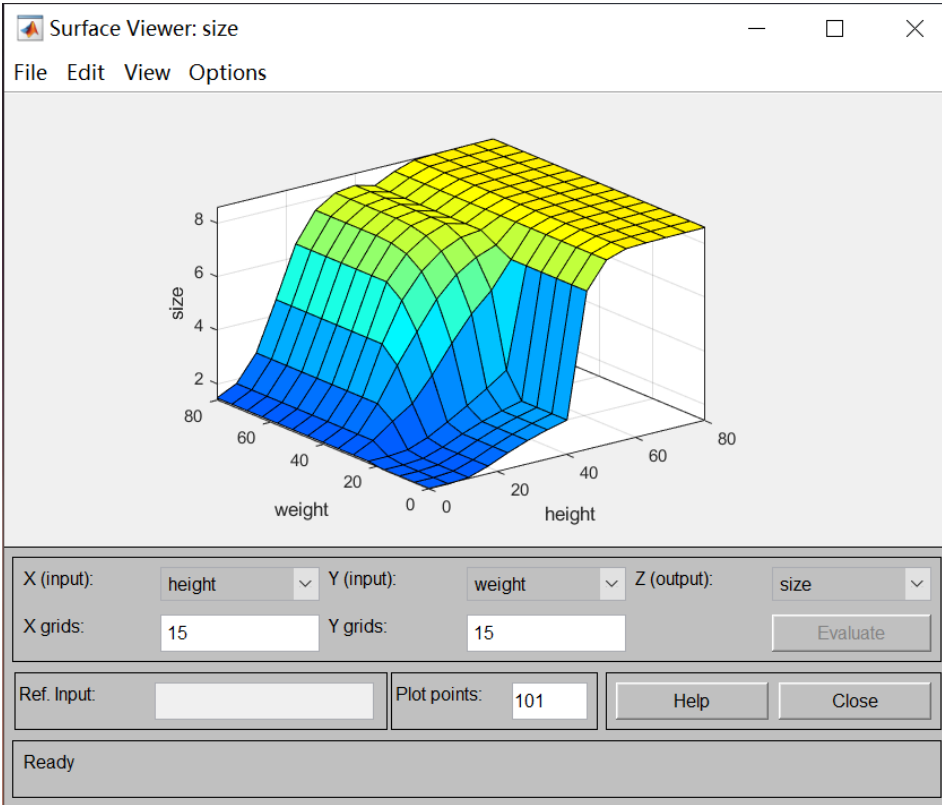
centroid

- 用规则观测器计算输出  
以金毛犬为例计算其体型

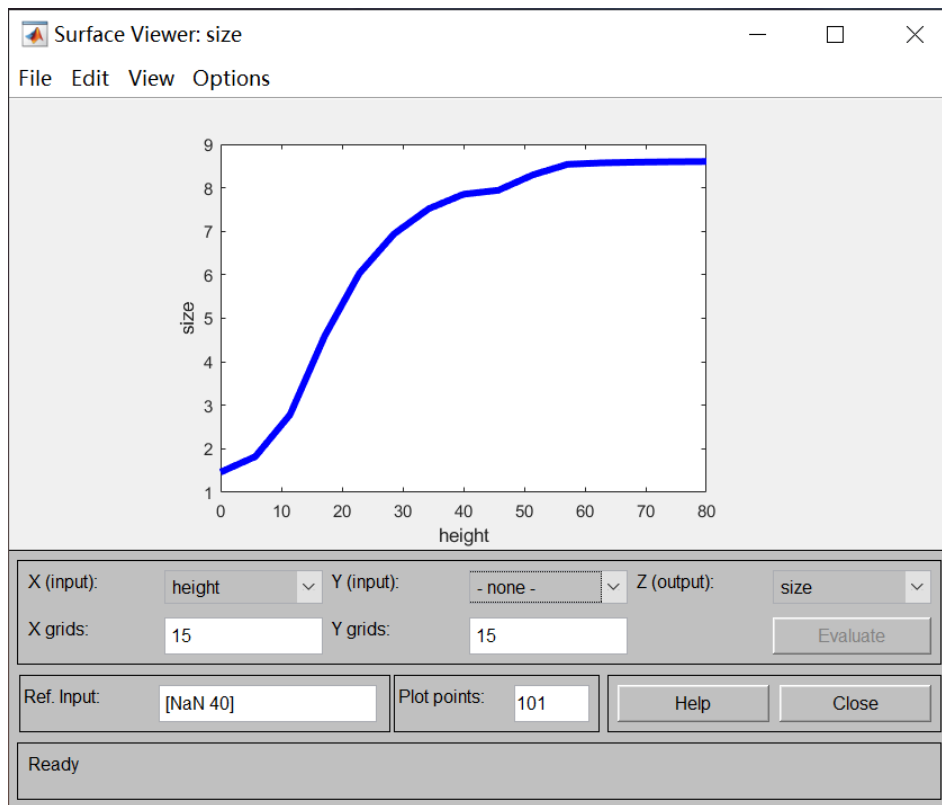


可以看到其使用规则进行推理的过程，最终结果金毛体型大小为 8.49/10

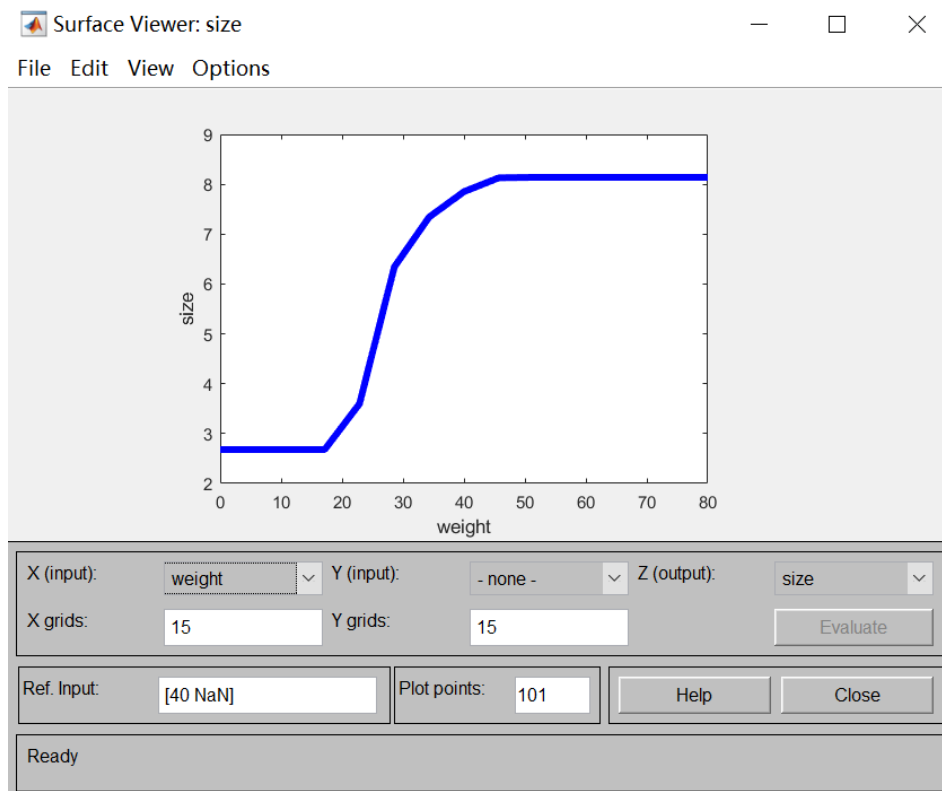
- 用曲面观测器查看依赖关系图



也可以只看体型-身高关系图

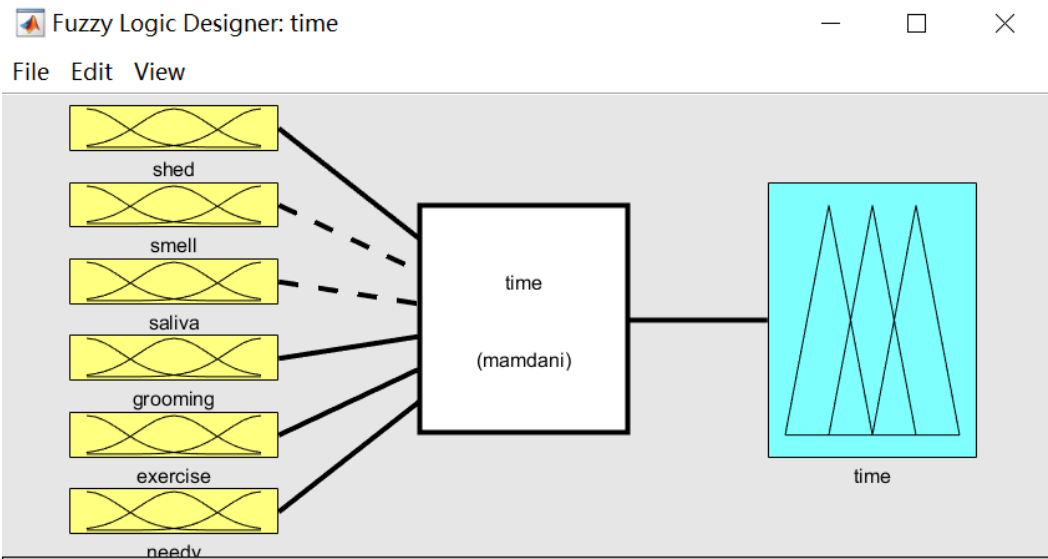


或者体型-体重关系图

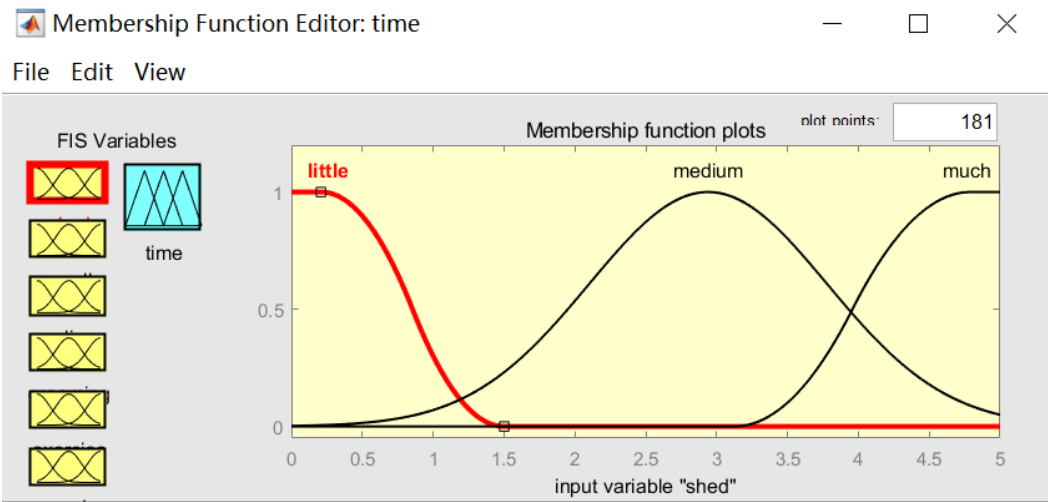


可以看到由于规则的内容，身高对体型的影响比较大，身高高一定体型大，在身高中等的时候体重重才可得出体型大，在身高较矮时体重对体型贡献不大

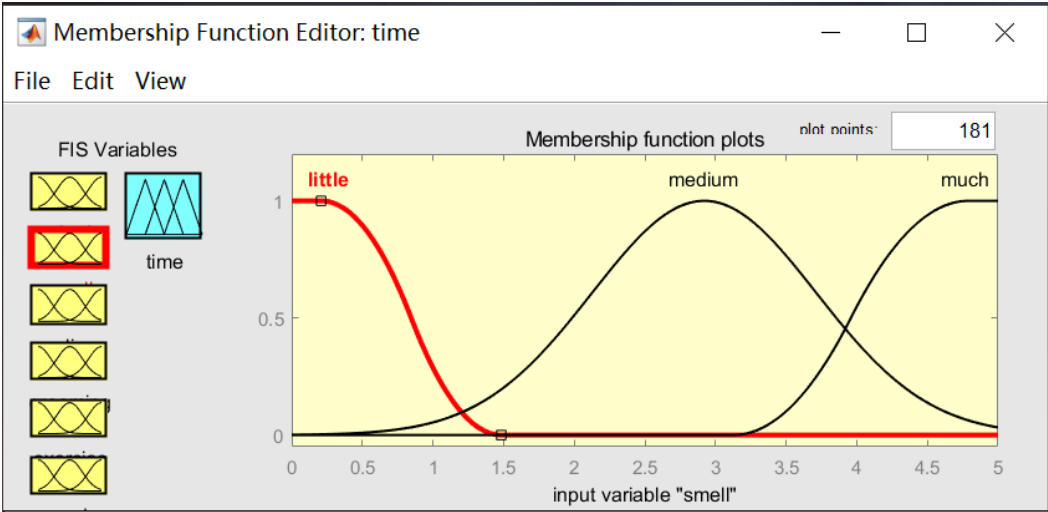
同样可以得到时间花费的模糊推理



掉毛程度 (shed)

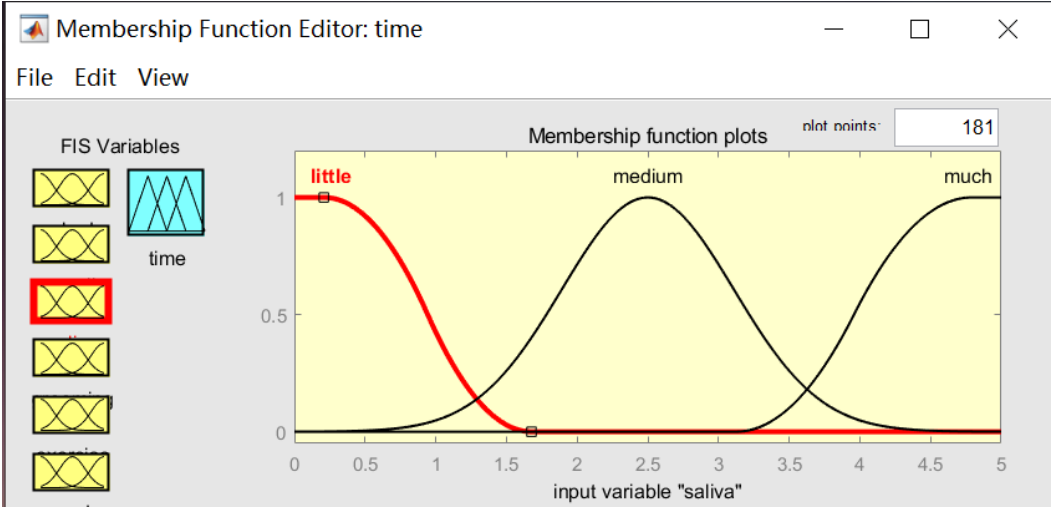


体味程度 (smell)

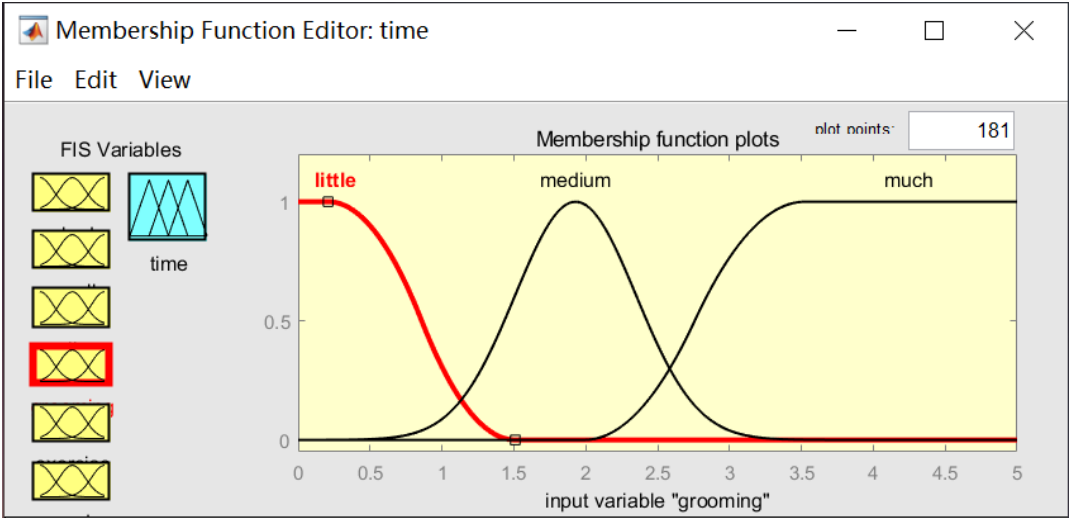




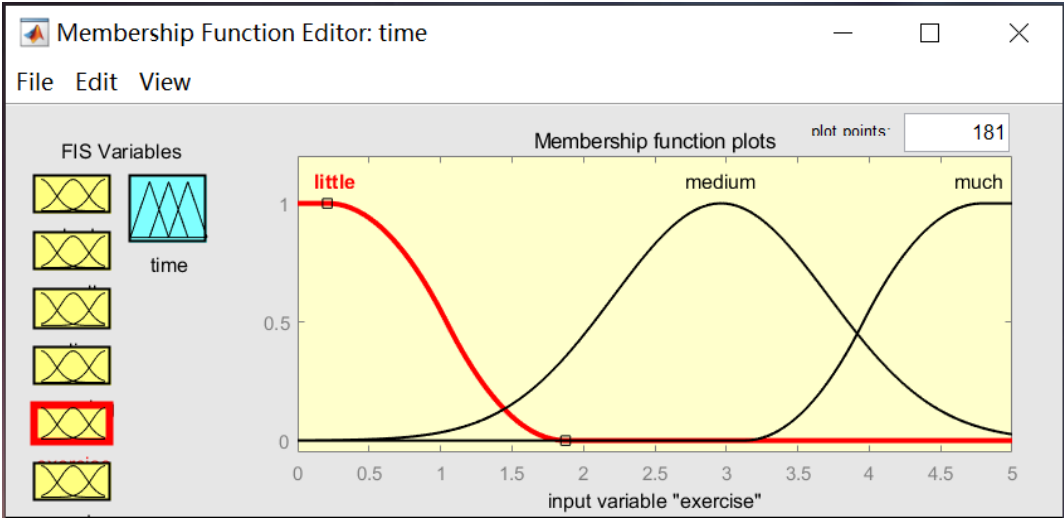
口水程度(saliva)



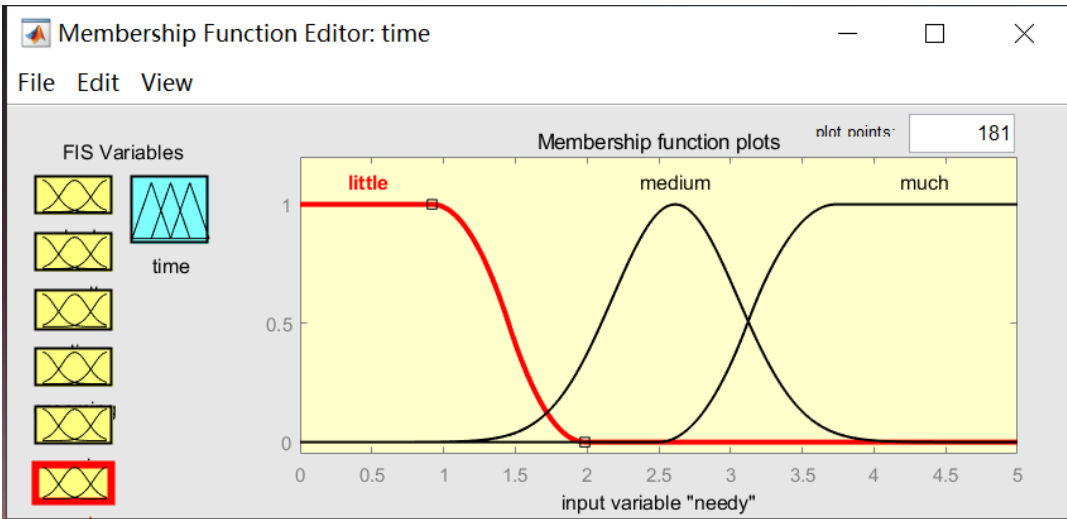
美容程度 (grooming)



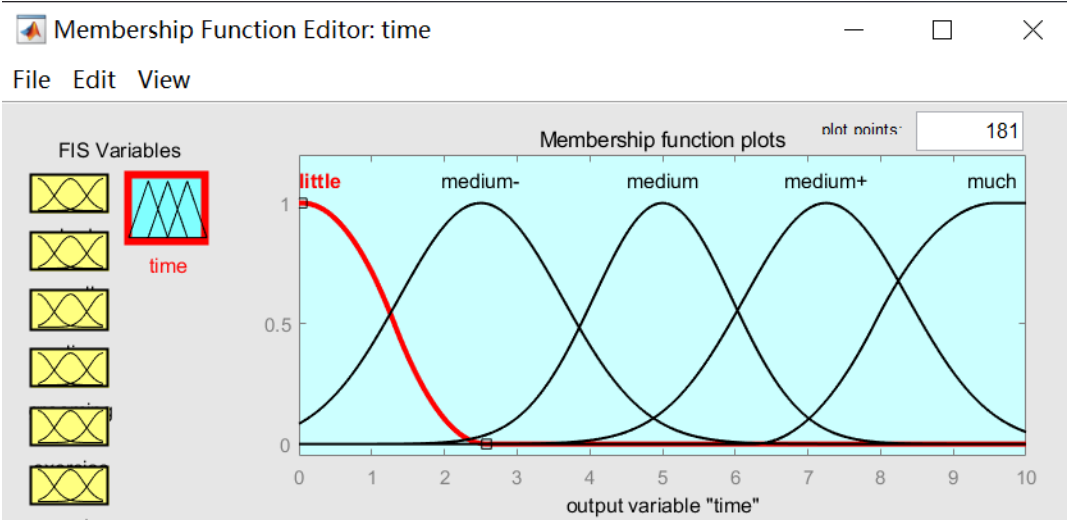
运动程度 (exercise)



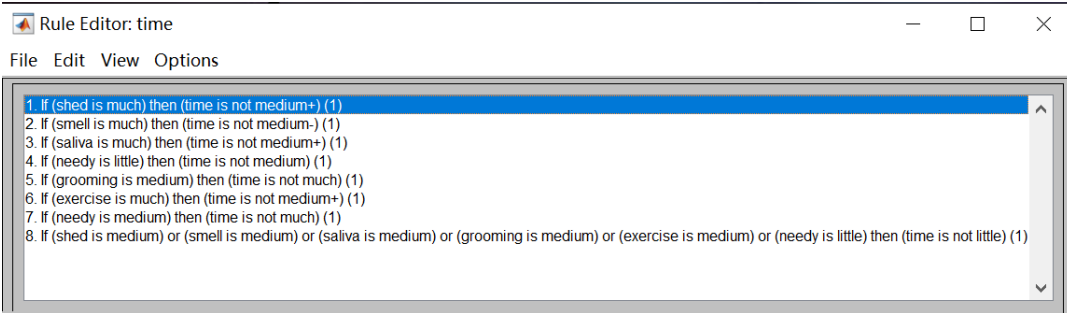
关爱需求度(needy)



打理照顾时间



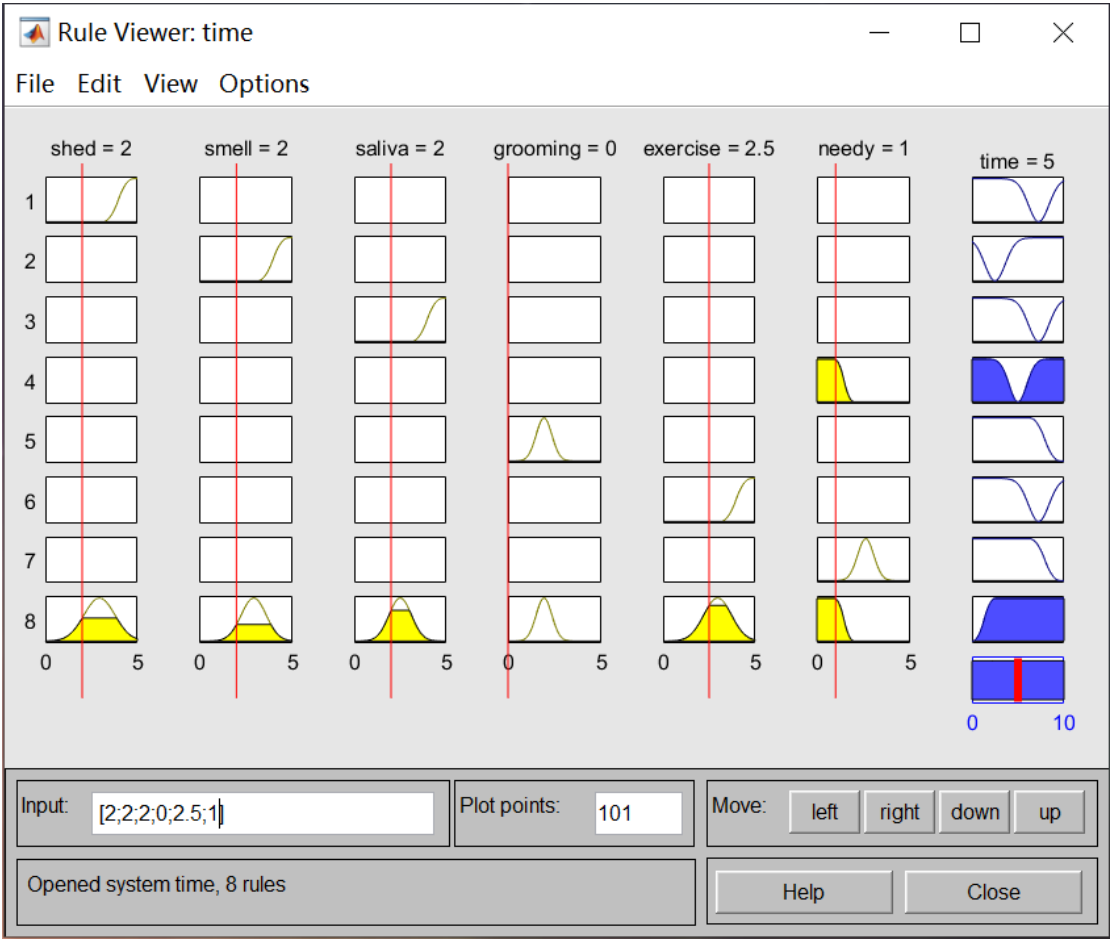
规则



后向使用 not 是为了方便最后的评分，即打理时间越多，该项得分越低

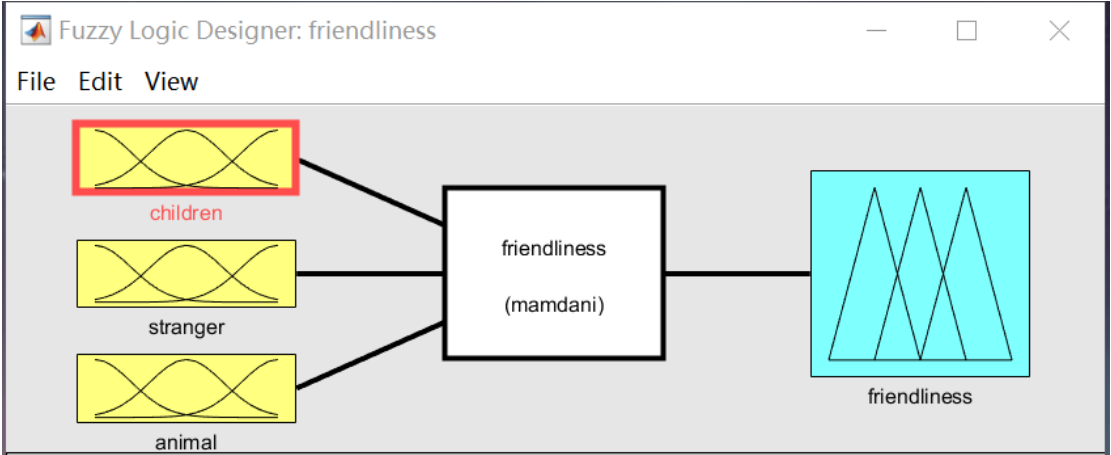
这里调整参数花了较多精力，主要是根据日常经验来修改。例如掉毛、体味、口水程度、运动程度很高时，才会认为花费时间多，因为大部分狗都有这些问题；而美容程度、关爱需求度中等时，就会花费很多时间，例如一些长毛犬需要日常维护，小型犬通常比较粘人

以中华田园犬为例计算输出

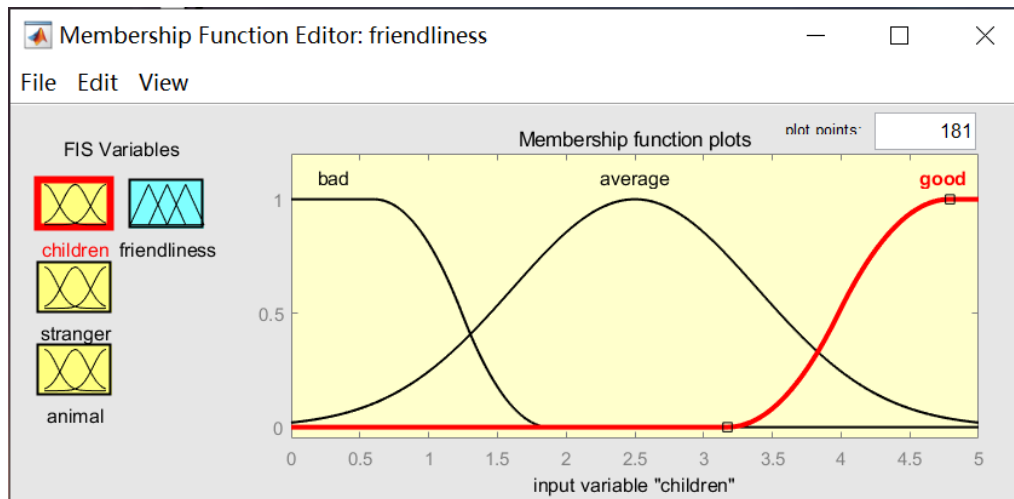


虽然田园犬较为独立，通常认为比较好养，但是因为口水、体味等问题，只得到平均分，当然这也有数据准确性的问题

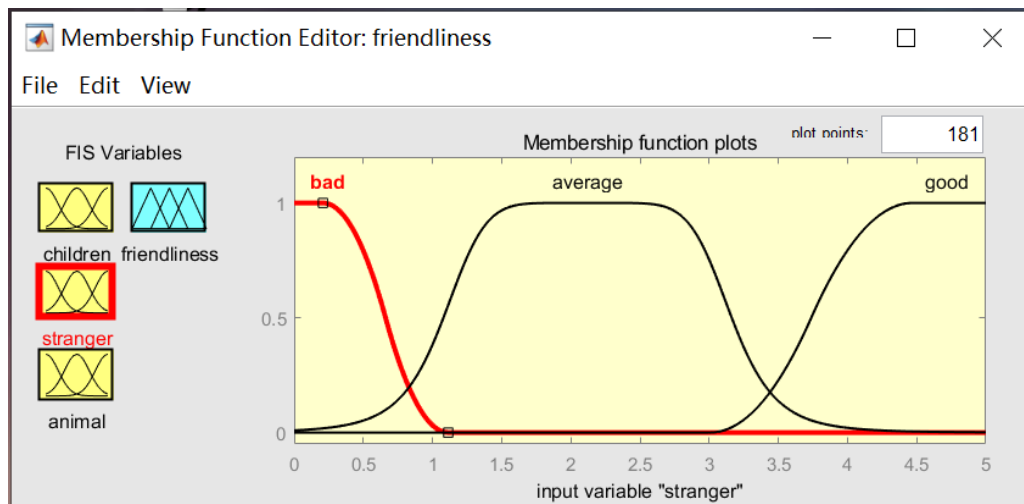
友善度的模糊推理同理



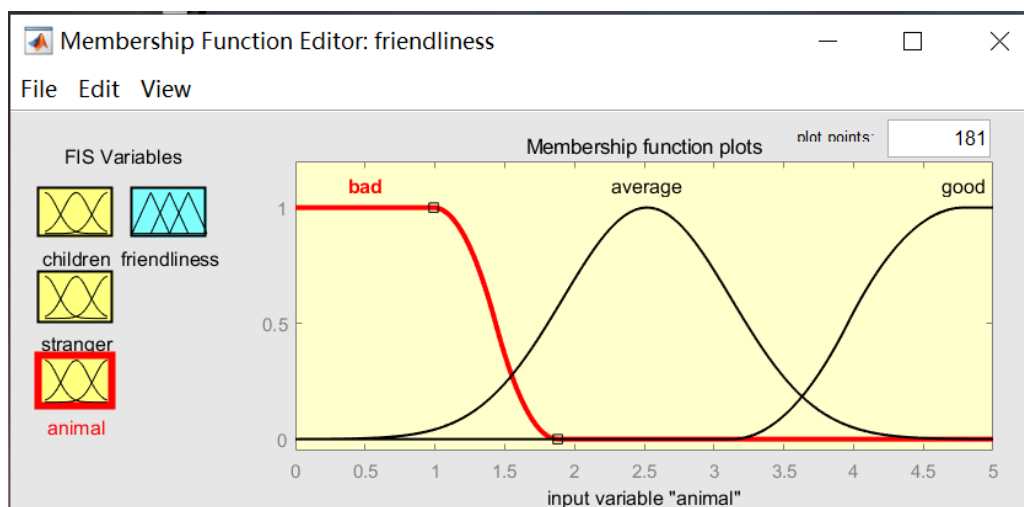
## 对小孩友善度



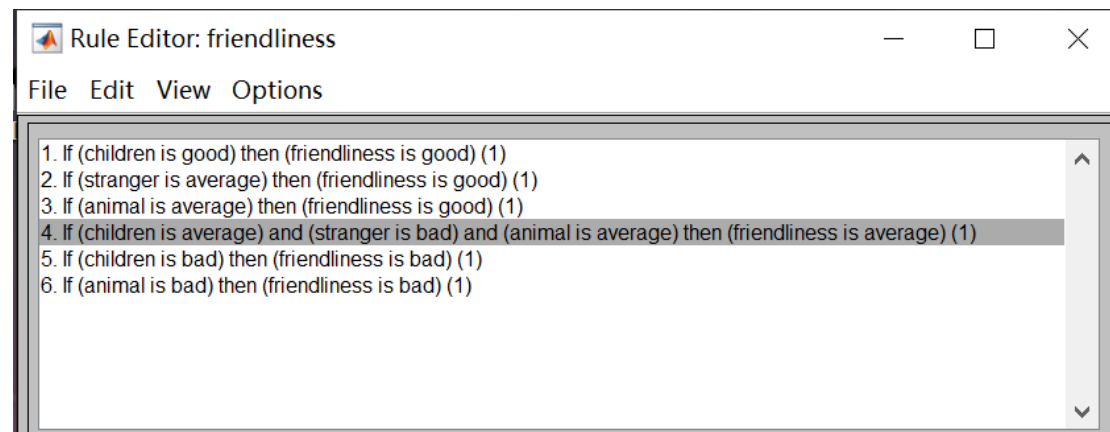
## 对陌生人友善度



## 对其它动物友善度

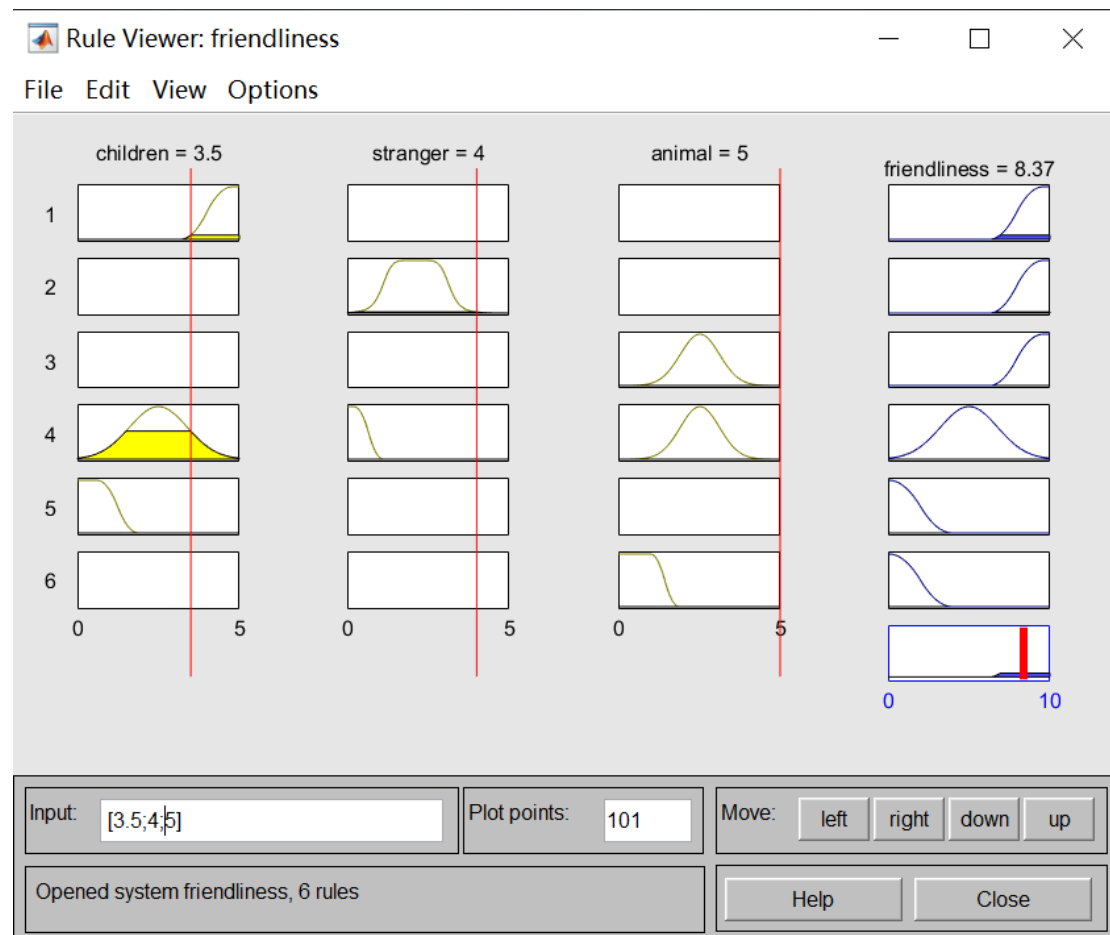


## 规则



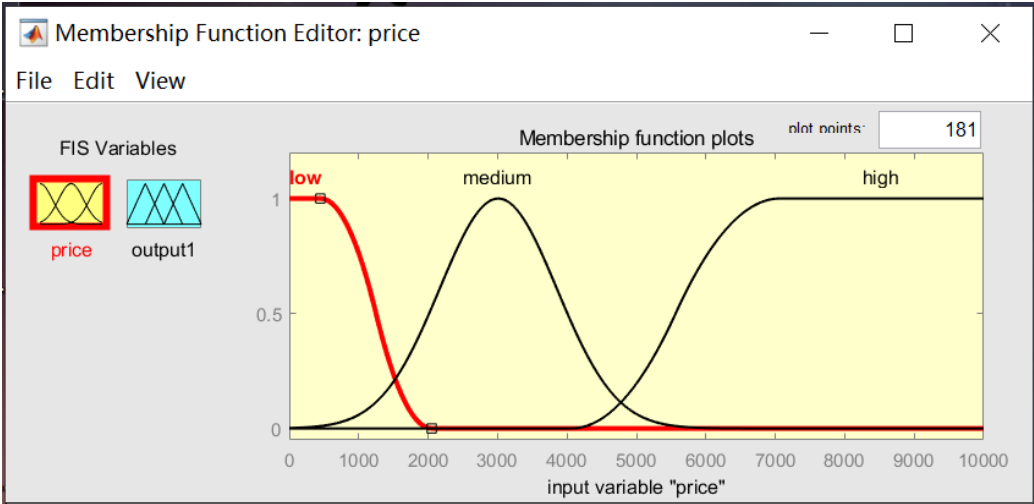
我们认为对于小孩和其它动物友善度较高则总体友善度高, 而对陌生人友善度低恰恰是狗的特点, 或者说是主人所需要的

以萨摩耶为例

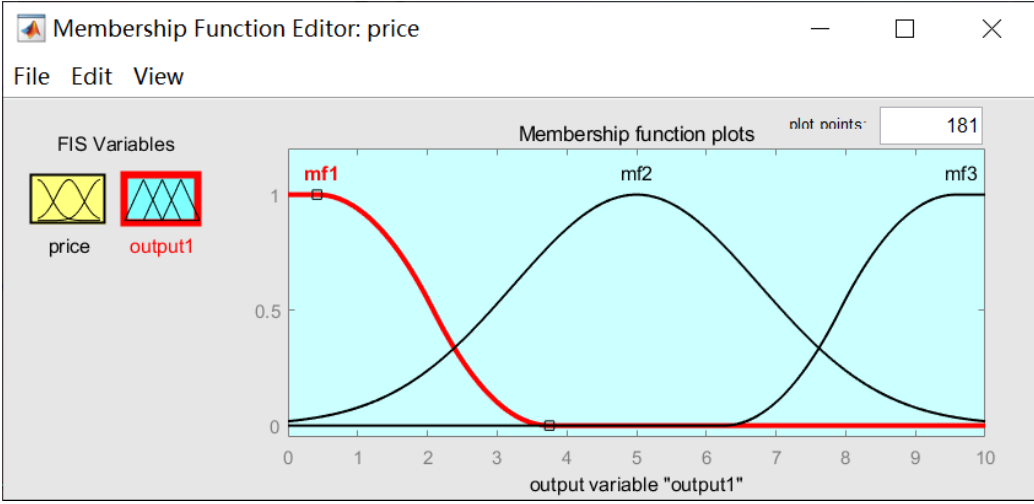


可以看出萨摩耶友善度很高

以及因为网页给出的是具体的**价格**，所以需要对应到 10 分制得分输入

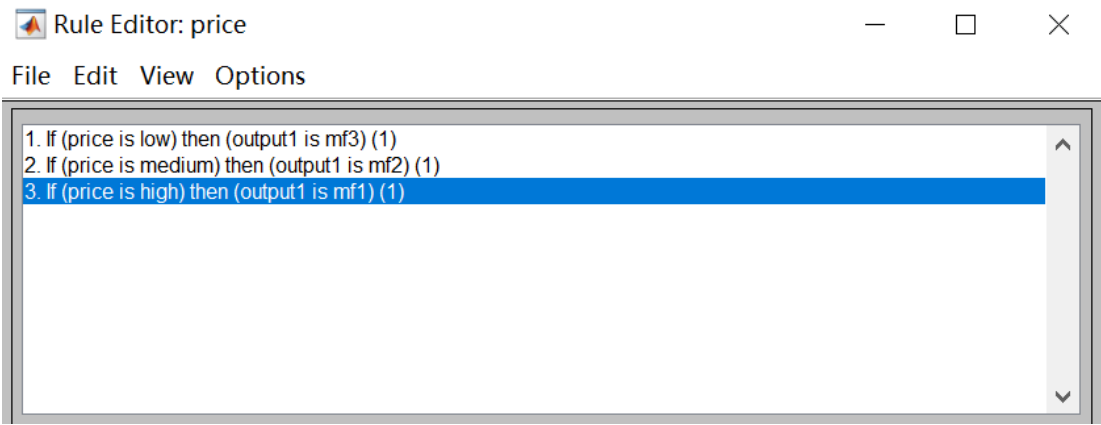


输出

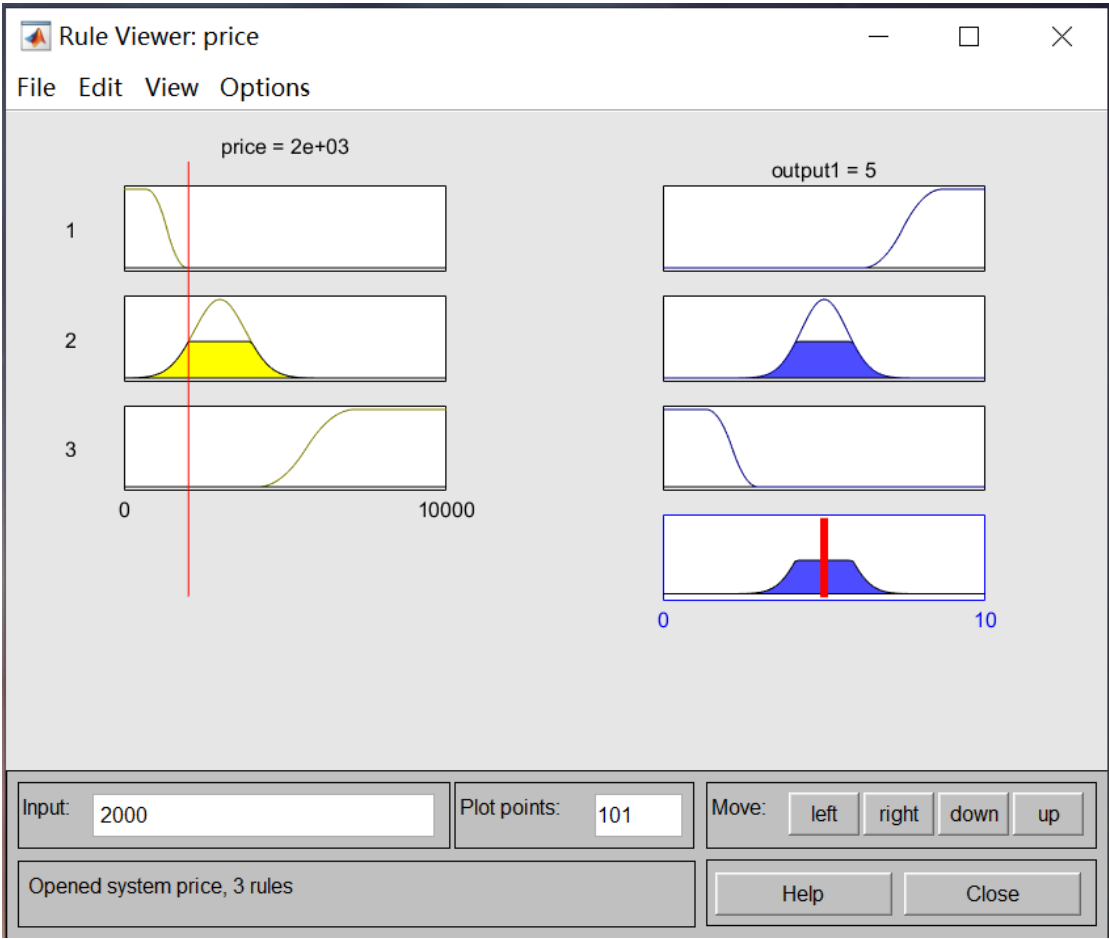


规则

需要注意价格越高，得分越低



以泰迪（售价大约 2000）为例，输出为 5 分



### ● 打分实现

获得数据和规则后，可以开始模糊推理。编写 C++ 程序完成前端的设计

通过前面提到的规则观测器，**手动输入数据**，得到体型、时间、友好度的相应分值，再加上现成的数据（例如可训练度、犬叫程度），**每个种类各单项得分**得到一个 11\*10 的表格

```
4 double value[11][10] = {
5     4.98, 5, 6.24, 7, 8, 10, 4, 3, 7, 8.86,
6     8.16, 4.23, 3.89, 4, 6, 3, 2, 3, 8, 5,
7     7.75, 4.68, 5.93, 8, 3, 3, 5, 7, 9, 5,
8     7.51, 6.54, 8.54, 5, 1, 7, 8, 10, 7, 5,
9     7.89, 6.65, 8.19, 8, 1, 3, 6, 10, 7, 5,
10    1.51, 4.68, 8.64, 10, 6, 4, 8, 10, 5, 5,
11    1.51, 6.65, 5.97, 6, 6, 1, 4, 6, 10, 5,
12    1.42, 3.78, 8.54, 8, 8, 10, 6, 10, 5, 1.17,
13    1.45, 4.29, 8.54, 10, 6, 10, 8, 7, 5, 5,
14    1.85, 4.16, 8.37, 4, 5, 1, 4, 8, 10, 5,
15    1.41, 3.71, 6.24, 8, 10, 5, 4, 6, 3, 1.13
16 };
```

接下来需要把用户的需求和适合度进行整合，得到最后的总评分（满分 100 分）

这里由用户对各指标的关心程度进行排序，按次序对应不同的权重，例如指标 0-体型排序序号为 3，即为第四关心（从 0 开始）的量，则狗的体型单项评分需要计算  $value \times (10 - 3) / 10$ ，即乘 0.7，所有十个指标的值相加得到最后的 score，对十一种宠物狗品种的最后 score 进行排序，取最大值对应的狗，输出其信息

四、用户界面及运行效果

已经将程序打包成 exe 文件，直接运行即可



dog.exe

首先介绍该专家系统

```
*****宠物狗推荐专家系统*****
说明：
0) 本系统仅作参考
1) 本系统暂时包含11种类型宠物狗：
    Number      Dog
    -----
    0           中华田园犬
    1           腊肠
    2           柯基
    3           泰迪
    4           比熊
    5           金毛
    6           哈士奇
    7           伯恩山
    8           拉布拉多
    9           萨摩耶
    10          藏獒
2) 本系统暂时包含10个参数：
    Number      Para
    -----
    0           体型小
    1           打理照顾时间短（包括掉毛、体味、口水、美容、运动、依赖性等）
    2           友善度（对小孩、生人、动物）高
    3           可训练度高
    4           忠诚度及护卫能力高
    5           初养适应度高
    6           犬叫程度低
    7           颜值高
    8           搞笑程度高（参照二哈）
    9           价格低
```

关心程度排序为：**价格低**>体型小>打理照顾时间短>友善度高>可训练度高>忠诚护卫能力强>初养适应能力高>犬叫程度低>颜值高>搞笑程度高>价格低时，系统推荐”中华田园犬”

```
请按参数关心程度排序：
1. 9
2. 0
3. 1
4. 2
5. 3
6. 4
7. 5
8. 6
9. 7
10. 8

最推荐您养的狗狗序号是： 0
各项参数评分为：
    Para      Score
    -----
    体型小      4.98
    打理照顾时间短      5
    友善度高      6.24
    可训练度高      7
    忠诚度及护卫能力高      8
    初养适应度高      10
    犬叫程度低      4
    颜值高      3
    搞笑程度高      7
    价格低      8.86

请按任意键继续. . .
```



当相对需要**体型小且易打理**时，系统推荐**柯基**

```
请按参数关心程度排序：
1. 0
2. 1
3. 2
4. 9
5. 3
6. 4
7. 5
8. 6
9. 7
10. 8

最推荐您养的狗狗序号是： 3
各项参数评分为：
      Para                Score
      -----
      体型小                7.51
      打理照顾时间短        6.54
      友善度高              8.54
      可训练度高            5
      忠诚度及护卫能力高    1
      初养适应度高          7
      犬叫程度低            8
      颜值高                 10
      搞笑程度高            7
      价格低                 5

请按任意键继续. . .
```

当需要**可训练程度高且忠诚、护卫能力强**时，系统推荐**拉布拉多**

```
请按参数关心程度排序：
1. 3
2. 4
3. 5
4. 6
5. 7
6. 8
7. 9
8. 0
9. 1
10. 2

最推荐您养的狗狗序号是： 8
各项参数评分为：
      Para                Score
      -----
      体型小                1.45
      打理照顾时间短        4.29
      友善度高              8.54
      可训练度高            10
      忠诚度及护卫能力高    6
      初养适应度高          10
      犬叫程度低            8
      颜值高                 7
      搞笑程度高            5
      价格低                 5

请按任意键继续. . .
```

其它实验结果包括：

当关心颜值时，小型犬会推荐泰迪、比熊；大型犬会推荐萨摩耶、金毛

当关心有趣搞笑程度时，小型犬会推荐柯基，大型犬会推荐哈士奇

当关心忠诚度及护卫能力，不在乎价格时，会推荐藏獒

.....

## 五、实验中遇到的问题、不足之处及实验感悟

本宠物狗推荐系统基本实现了初始的构想，能够针对用户的特点匹配出最佳宠物狗种类。在实验过程中主要遇到了如下已解决和未解决的问题：

### 1) 参数调整

“开发模糊专家系统，最费力和枯燥的步骤是调试。”本次实验中深深体会到了这一点。

前面三个指标主要需要调整隶属度函数参数、变量范围。一开始用的三角隶属函数，均值默认，评估结果与期望大不相同，甚至相反，期望差别较大的结果，例如比熊的体型和萨摩耶体型相近…；又例如一开始没有取反（价格高则推荐度低），导致结果不符合常识

虽然经过调整，最终结果较为合理，但是还有很大改进空间，例如模糊规则可以定义每条规则的权重（weight）（默认为 1）；最后汇总的时候用户排序时，真实情况可能是只关心几个最前面的几个指标，或者一些指标是可以同等重要的，所以乘的比重还需要进一步地实验、评估、调整，可以做成用户自己选择比重

### 2) 数据不准确、数据量不足

10 项指标已经较为全面，但仍然可以补充完善，例如寻回、导盲等功能

狗的品种过于少，这一点是因为 matlab 模糊逻辑无法导入 excel 表格，不能批量输入和输出，所以只能手动输入计算，再把结果放到表格里，所以只选择了总共大约 300 类狗中差异较为明显的、较为熟悉的 11 个品种，这样就导致分析结果可能并不是 300 种中最适合的。网络上数据存在很多问题，比如精度不够，可靠性低，使得宠物狗推荐系统可能不如其它推荐系统（如电脑手机推荐有准确的参数）准确。

进一步解决该问题，考虑爬取上面列出的两个网站的数据，换用其它模糊逻辑语言编辑器，需要去学习除了 matlab 以外新的工具！

### 3) 用户界面不够友好

在做该宠物狗推荐系统之前先做过一个简单的**白葡萄酒识别系统**（后来因为觉得过于简单且没有模糊逻辑而放弃），如下：



使用了 java GUI 界面，但是因为没有学习如何把模糊逻辑嵌入到 java 中，所以在做后来的推荐系统时只能采取这种很笨拙的方式，通过 matlab 计算，然后用 C++ 统计输出，命令行作为界面。

在本次实验中，我充分体验了设计专家系统的全过程，包括主题的确定（一个小而精的领域）、采集数据（知识库的建立）、模糊集的定义、推理规则书写，到最后整合输出并设计用户交互界面。虽然最后的结果不够完善，还有更为方便美观的工具去学习，但是基本完成了初期的想法，也体会到了专家系统的意义所在，收获很多。但感到遗憾和愧疚的是，虽然很早开始构想感兴趣主题，但是没有很早付诸实践，导致期末季才快速完成，所以说还有很多很多需要改进的地方！