

基于模糊专家系统的香水推荐系统

陈俊含 19307130180 2022.06.27

一、项目基本信息

1. 性质：模糊专家系统
2. 推理引擎：Fuzzy Logic Toolbox
3. 前后端：Django + HTML
4. 数据库：sqlite3
5. 操作系统：Win10
6. IDE：VSCode
7. 代码地址：[Github](#)

二、工程文件架构（只列举核心文件）

1. backend/
 - backend/
 - src/ : 存放 html 文件，即用户界面
 - static/ : 存放所有香水的图片
 - perfumes/
 - fuzzyLogic/ : 存放需要使用用户输入信息的模糊规则文件
 - forms.py : 前端供用户填写的表单
 - getData.py : 即爬虫，用于爬取所有的香水数据并存入数据库
 - models.py : 指定数据库中存放的表及其所有属性信息
 - readyForRate.py : 预先填写一些信息以便后续的评分
 - preRate.py : 先使用不需要用到用户输入信息的模糊规则进行评分，并写入数据库
 - style.m 和 year_score.m : matlab 脚本，用于启动模糊推理引擎并返回计算结果给后端
 - db.sqlite3 : 数据库
 - toExcel.py : 将数据库中的信息转化为 excel 表格，便于模糊推理引擎批量计算数据
2. fuzzyLogic/ : 存放不需要使用用户输入信息的模糊规则文件，可预先使用
3. pics/ : 存放本报告所需的图片
4. xlsx/ : 存放导出的一些 excel 表，包括模糊推理引擎所需的输入和最终输出

三、主要建构过程

1. 新建 Django 项目，配置好静态文件目录、模板目录等构件
2. 新建名为 perfumes 的 app，设计数据库的两个表（属性等信息），存储后续爬取或计算得到的数据
3. 爬取香水数据

- 我选择的主站点为香水时代，爬取了其 TOP200 榜单的 200 款香水的详细数据，包括名称、图片、香调类型、前中后调、留香时间等
 - 在爬取过程中自动将数据写入数据库中，供后续打分、查询数据时使用
 - 爬取的数据存储在名为 Perfume 的表中
- 4.将 Perfume 表中的与评分有关的信息填写进 Rate_Perfume 表中，如留香时间、香调种类等
- 5.预评分阶段
- 将 Rate_Perfume 表中的 留香时间time、排名rank、用户评分rate 三列数据预先计算 留香得分 time_score、用户喜爱度评分user_love_rate 两个指标值
 - 具体方式为将上述提到的三列数据导出到 excel 文件，在matlab工作台中调用 lasting_time 和 user_love 两个模糊规则进行计算，将得到的结果也写入新的 excel 表中，最后将表中两列的数据写入 Rate_Perfume 中对应的属性中
- 6.前端构建
- 即编写前端页面（html文件）和供用户填写的表单格式，并在 views 模块中编写对应的函数以处理用户在前端填写的数据
 - 处理包括：检查表单、从表单中获取数据、将用户数据和 Rate_Perfume 中的数据作为输入调用 style.m 和 year_score.m 两个模糊规则脚本、将得到的结果写回数据库中、计算所有符合标准的香水的最终推荐度、根据推荐度降序排列并取前七个、跳转到推荐结果的页面

四、使用方式

- 1.cmd 定位到第一个 backend 文件夹，命令行输入 `python manage.py runserver`
- 2.在浏览器打开 `http://127.0.0.1:8000/`，页面如下

← → ↻ 127.0.0.1:8000

香水推荐

男香:	<input type="checkbox"/>	不考虑它则勾选
中性香:	<input type="checkbox"/>	不考虑它则勾选
女香:	<input type="checkbox"/>	不考虑它则勾选
用户喜爱度偏好:		范围为[0,1] 数值越大在乎程度越大
留香时间偏好:		范围为[0,1] 数值越大留香时间越长
年代偏好:		范围为[0,1] 数值越大则更偏好新款
清新偏好:		范围为[0,1] 数值越大喜爱程度越大
古典偏好:		范围为[0,1] 数值越大喜爱程度越大
浓郁偏好:		范围为[0,1] 数值越大喜爱程度越大
芬芳偏好:		范围为[0,1] 数值越大喜爱程度越大

点击推荐

Copyright @Brayton

- 3.根据提示和要求填写表单，点击 `点击推荐` 按钮，等待几秒钟
- 4.得到7款推荐的香水，包含每款香水的详细信息，效果如下图

香水推荐												
推荐度排名	推荐度	名称	图片	品牌	香调	前调	中调	尾调	属性	调查师	标签	最明显的气味
1	0.5237104381153828	迪奥 华氏温度 Dior Fahrenheit, 1988		迪奥	馥奇香调	薰衣草 橘子 山楂 肉豆蔻 雪松 香柠檬 洋甘菊 柠檬	肉豆蔻 忍冬 康乃馨 檀香木 紫罗兰叶 茉莉 铃兰 雪松	皮革 零陵香豆 琥珀 广藿香 麝香 香根草	男香	Jean-Louis Sieuzac	皮革 木香 水生 清新辛辣 动物 浓郁 水感 花香 柑橘 绿叶	皮革 紫罗兰叶 薰衣草 肉豆蔻 香根草
2	0.5056875810525964	阿莎罗 同名男士 (卡门情人) Azzaro Pour Homme, 1978		阿莎罗	馥奇香调	葛缕子 鸢尾花 薰衣草 快乐鼠尾草 罗勒 茴芹 香柠檬 柠檬	檀香木 杜松子 广藿香 香根草 雪松 小豆蔻	皮革 零陵香豆 琥珀 麝香 橡木苔	男香	Martin Heidenreich	浓郁 木香 清新辛辣 泥土 薰衣草 茴芹 柔和辛辣 苔藓 皮革 柑橘	薰衣草 茴芹 柠檬 橡木苔 皮革
3	0.5047766837513928	博柏利 周末男士 Burberry Weekend for Men, 1997		博柏利	柑橘调	菠萝 香瓜 橘子 葡萄柚 香柠檬 柠檬	檀香木 常春藤 橡木苔	蜂蜜 琥珀 麝香	男香	Michel Almairac	柑橘 清新 甜 果香 浓郁 绿叶 水生 木香	柠檬 橘子 香柠檬 葡萄柚 麝香
Copyright ©Brayton												

香水推荐												
4	0.5031745999743406	切瑞蒂 1881男士 Cerruti 1881 Men, 1990		切瑞蒂	馥奇香调	薰衣草 康乃馨 白松香 香柠檬 杜松 柏树 桉香脂	香根草 依兰 玫瑰 铃兰 醋栗叶	麝香 檀香木 广藿香 橡木苔 雪松 胡椒 松树	男香	Martin Gras	木香 浓郁 清新辛辣 泥土 绿叶 苔藓	松树 薰衣草 香柠檬 橡木苔 杜松
5	0.5001694091496165	香奈儿 绅士 Chanel Pour Monsieur, 1955		香奈儿	西普调	橙花油 苦橙叶 柠檬	芫荽 生姜 罗勒 小豆蔻	橡木苔 香根草 雪松	男香	Henri Robert	柑橘 浓郁 木香 泥土 苔藓 温暖辛辣 清新 绿叶 清新辛辣 草本	柠檬 橡木苔 生姜 香根草 柠檬 马鞭草
6	0.4916528593581858	迪奥 旷野 Dior Sauvage, 2015		迪奥	馥奇香调	胡椒 香柠檬	天竺葵 薰衣草 花椒 桉香脂 粉红胡椒 香根草 广藿香	雪松 劳丹脂 降龙涎香醚	男香	Francois Demachy	清新辛辣 琥珀 柑橘 浓郁 麝香 木香 草本 温暖辛辣 薰衣草	降龙涎香醚 香柠檬 胡椒 薰衣草 雪松
7	0.47914944984888985	范思哲 同名经典男士 Versace Pour Homme, 2008		范思哲	馥奇香调	柠檬 橙花油 香柠檬 玫瑰	风信子 快乐鼠尾草 雪松 天竺葵	零陵香豆 麝香 琥珀	男香	Alberto Morillas	柑橘 浓郁 清新辛辣 花香 玫瑰 绿叶	柠檬 海水 橙花油 香柠檬 麝香
Copyright ©Brayton												

五、系统分析

(一) 指定问题并定义语言变量

1.问题：

- 引导用户填写个人喜好信息，定性（如想要的香水属性）与定量（如喜爱程度）相结合
- 避免使用香水相关的专业术语，而是用一些偏生活化的词语，增强其使用范围与可用性
- 根据得到的数据，利用模糊推理引擎计算数据库中符合条件的香水的推荐度，并根据推荐度降序排列，得到推荐度前七的香水品种并输出其具体信息

2.模糊规则中的语言变量

语言变量	变量性质	名词解释	注释
lasting_time	输入变量	香水的留香时间	数值越低代表其留香时间越短

语言变量	变量性质	名词解释	注释
time_score	输出变量	对于留香时间的评分	数值越高代表该款香水留香时间越长
rate	输入变量	用户评分	数值越大评分越高
rank	输入变量	排名	数值越小排名越高
user_love_score	输出变量	用户喜爱度评分	综合 rate 和 rank 两个指标，数值越大越受欢迎
year	输入变量	被研发出来的年份	
user_prefer	输入变量	用户对于年份的偏爱情况	数值越大代表越偏爱新款
year_score	输出变量	年份评分	综合 year 和 user_prefer 两个指标得出结果
清新、古典、浓郁、 芬芳	四个输入 变量	对于四个描述香调的形容词的 偏爱程度	数值越大越喜欢
柑橘调、绿叶调、水 生调	三个输出 变量	对于三种香调的喜爱度得分	这三个香调属于 清新 的范畴
馥奇香调、皮革调、 西普调	三个输出 变量	对于三种香调的喜爱度得分	这三个香调属于 古典 的范畴
木制调、东方调、美 食调	三个输出 变量	对于三种香调的喜爱度得分	这三个香调属于 浓郁 的范畴
花香调、果香调	两个输出 变量	对于两种香调的喜爱度得分	这两个香调属于 芬芳 的范畴

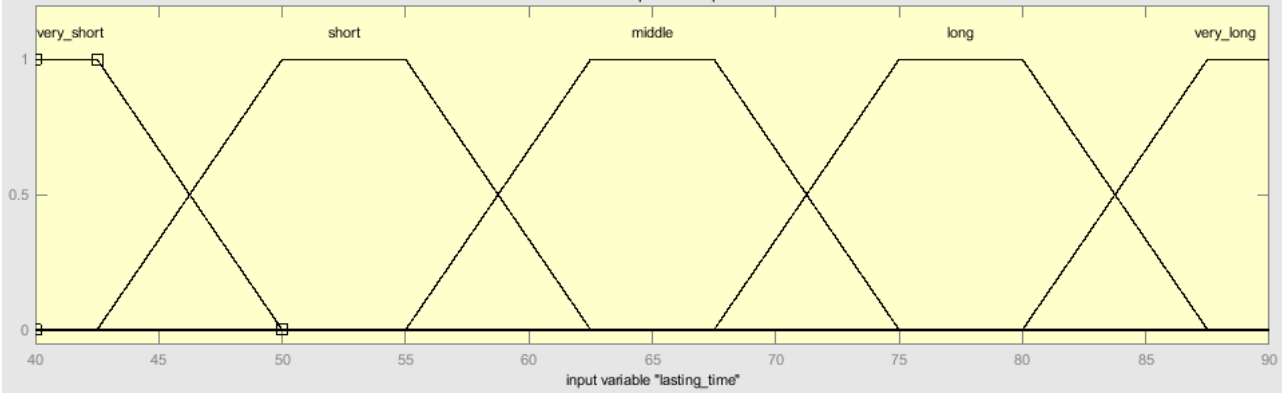
3.语言变量的范围（黑体变量为输出变量，使用高斯函数[标准差, 均值]）

语言变量	语言值1	语言值2	语言值3	语言值4	语言值5
lasting_time	very_short [40, 50]	short [42.5, 62.5]	middle [55, 75]	long [67.5, 87.5]	very_long [80, 90]
time_score	very_low [0.1062, 0]	low [0.1062, 0.25]	average [0.1062, 0.5]	high [0.1062, 0.75]	very_high [0.1062, 1]
rate	low [8, 8.75]	mid [8.25, 9.75]	high [9.25, 10]		
rank	very_high [1, 45.77]	high [5.975, 95.53]	middle [55.72, 145.3]	low [105.5, 195]	very_low [155.2, 200]
user_love_score	bad [0.1062, 0]	relatively_bad [0.1062, 0.25]	soso [0.1062, 0.5]	relatively_good [0.1062, 0.75]	good [0.1062, 1]
year	久远 [1860, 1896]	较久远 [1864, 1936]	中间 [1904, 1975]	较新 [1943, 2015]	很新 [1983, 2019]

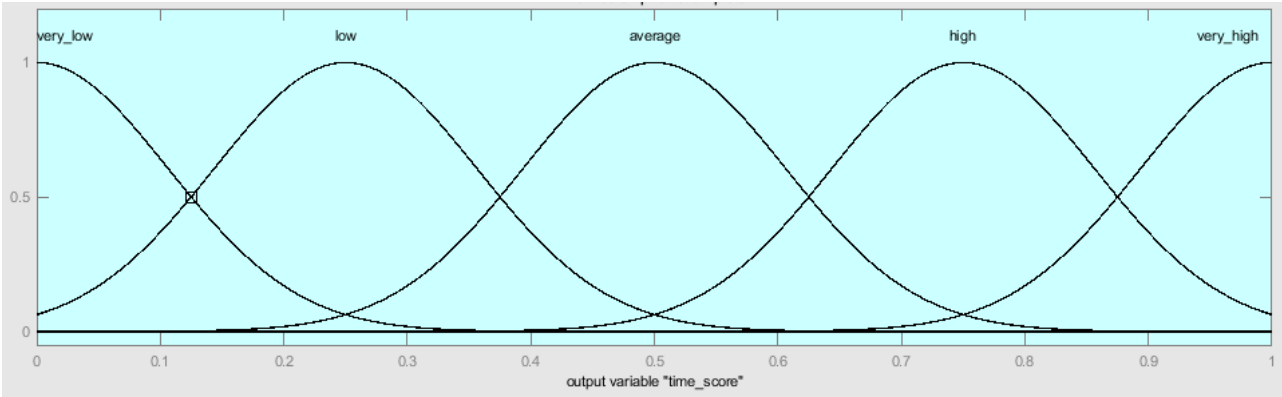
语言变量	语言值1	语言值2	语言值3	语言值4	语言值5
user_prefer	偏好久远 [0, 0.5]	偏好中间 [0.25, 0.75]	偏好新款 [0.5, 1]		
year_score	低 [0.1062, 0]	较低[0.1062, 0.25]	一般 [0.1062, 0.5]	较高 [0.1062, 0.75]	高 [0.1062, 1]
清新	讨厌 [0, 0.4]	无感 [0.2, 0.8]	喜欢 [0.6, 1]		
古典	讨厌 [0, 0.4]	无感 [0.2, 0.8]	喜欢 [0.6, 1]		
浓郁	讨厌 [0, 0.4]	无感 [0.2, 0.8]	喜欢 [0.6, 1]		
芬芳	讨厌 [0, 0.4]	无感 [0.2, 0.8]	喜欢 [0.6, 1]		
柑橘调	low [0.2123, 0]	mid [0.2123, 0.5]	high [0.2123, 1]		
绿叶调	low [0.2123, 0]	mid [0.2123, 0.5]	high [0.2123, 1]		
水生调	low [0.2123, 0]	mid [0.2123, 0.5]	high [0.2123, 1]		
馥奇香调	low [0.2123, 0]	mid [0.2123, 0.5]	high [0.2123, 1]		
皮革调	low [0.2123, 0]	mid [0.2123, 0.5]	high [0.2123, 1]		
西普调	low [0.2123, 0]	mid [0.2123, 0.5]	high [0.2123, 1]		
木质调	low [0.2123, 0]	mid [0.2123, 0.5]	high [0.2123, 1]		
东方调	low [0.2123, 0]	mid [0.2123, 0.5]	high [0.2123, 1]		
美食调	low [0.2123, 0]	mid [0.2123, 0.5]	high [0.2123, 1]		
花香调	low [0.2123, 0]	mid [0.2123, 0.5]	high [0.2123, 1]		
果香调	low [0.2123, 0]	mid [0.2123, 0.5]	high [0.2123, 1]		

(二) 定义模糊集

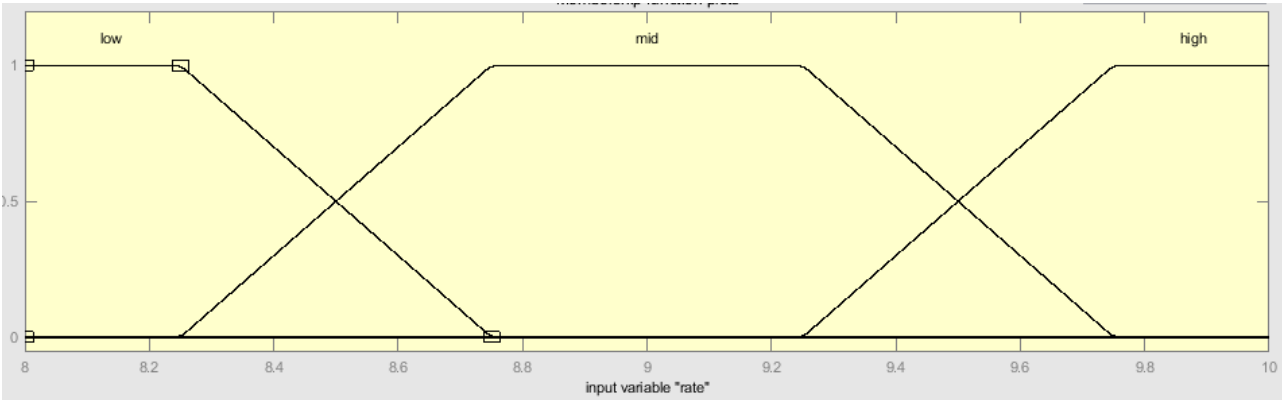
- 所有的输入变量采用三角形或四边形，而输出选用高斯函数
- 留香时间 lasting_time



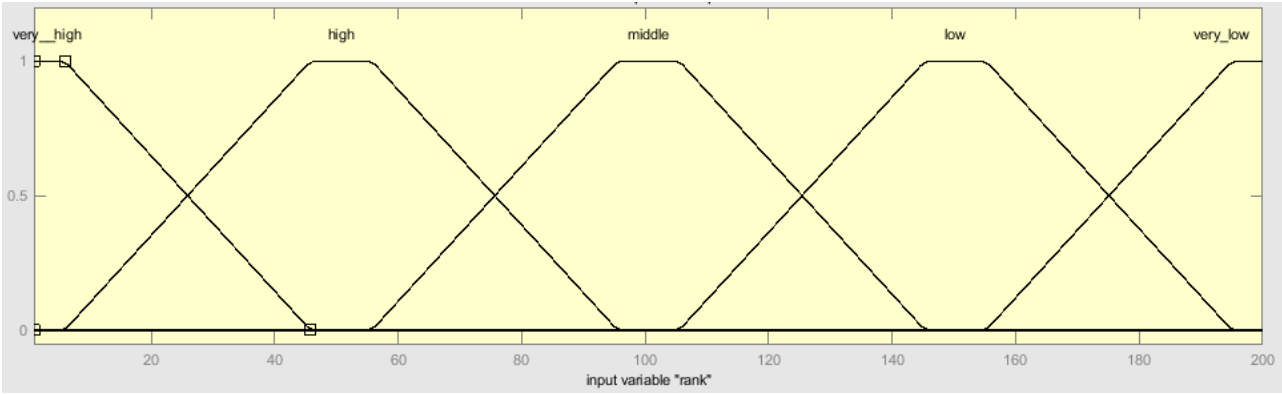
- 留香评分 time_score



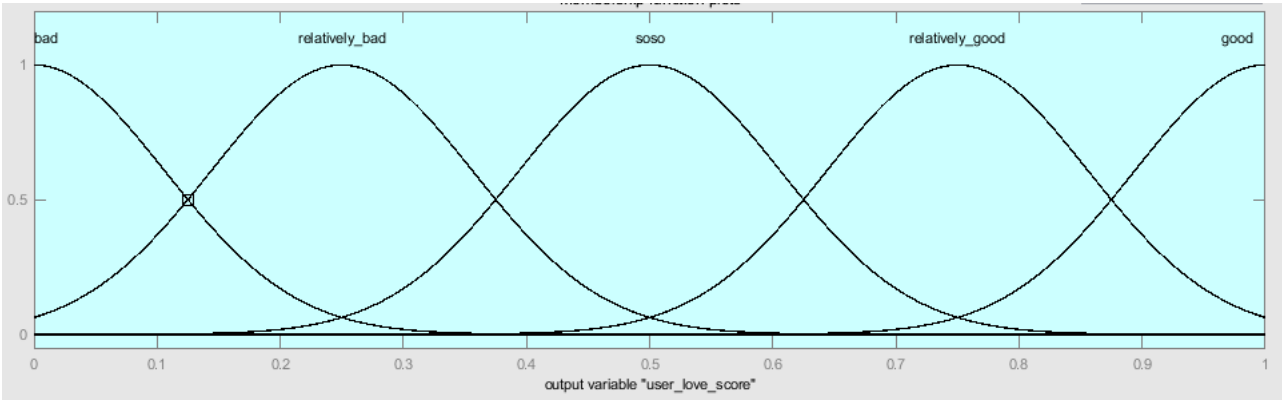
- 用户评分 rate



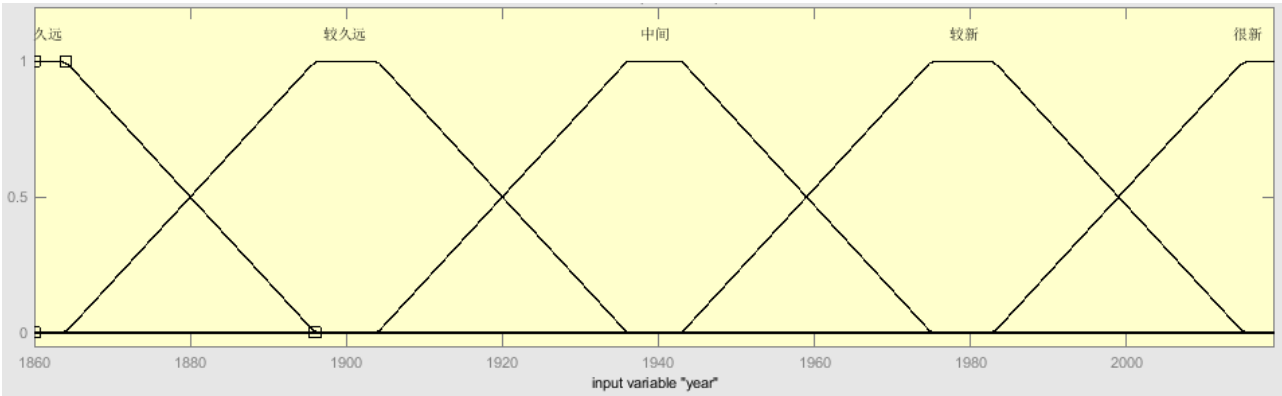
- 排名 rank



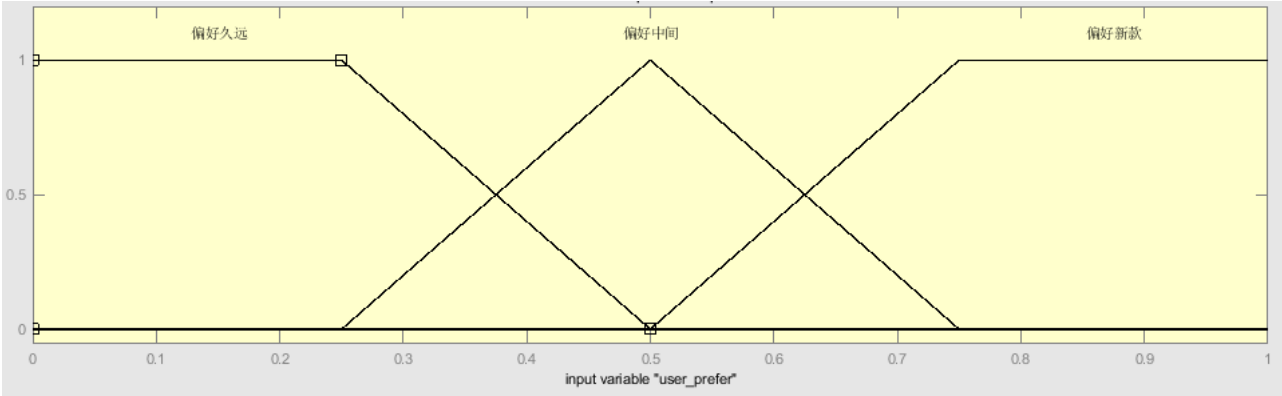
- 用户喜爱度评分 user_love_score



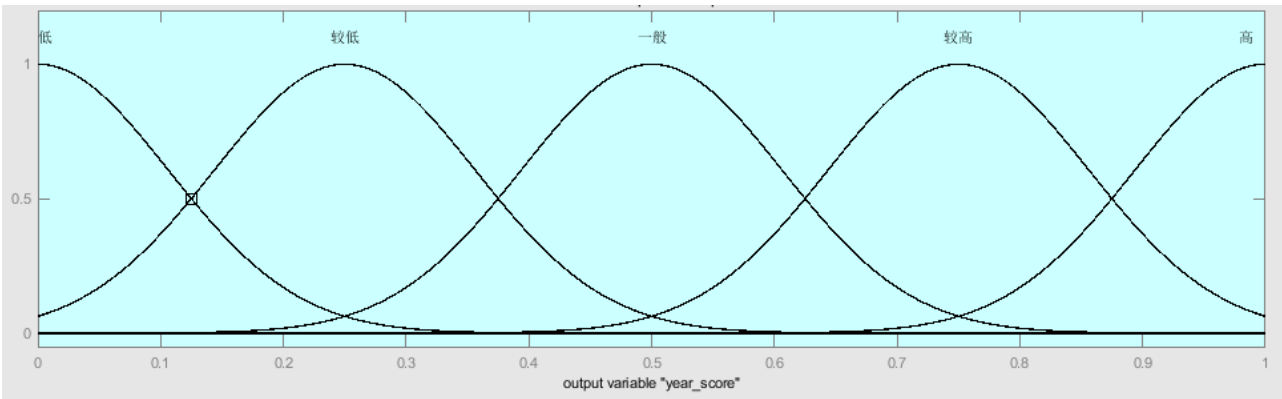
• 年代 year



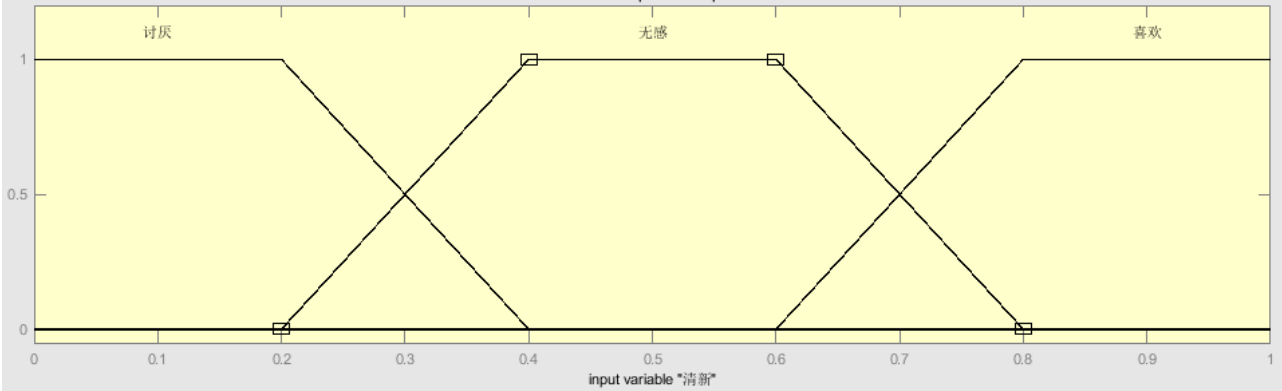
• 年代偏好 user_prefer



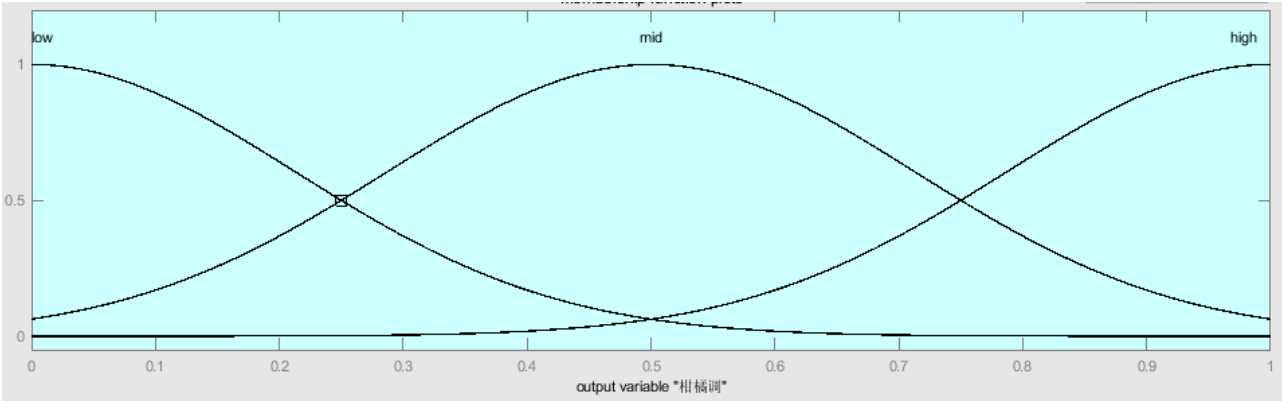
• 年代评分 year_score



• 清新（古典、浓郁、芬芳的模糊集也同下图）



- 柑橘调 (其它 10 种香调的模糊集也同下图)



(三) 抽取并构造模糊规则

Rule Base 1

- 用途：根据留香时间 lasting_time 计算留香评分 time_score
- 规则如下：

```
1. If (lasting_time is very_short) then (time_score is very_low) (1)
2. If (lasting_time is short) then (time_score is low) (1)
3. If (lasting_time is middle) then (time_score is average) (1)
4. If (lasting_time is long) then (time_score is high) (1)
5. If (lasting_time is very_long) then (time_score is very_high) (1)
```

Rule Base 2

- 用途：根据用户评分 rate 和排名 rank 计算用户喜爱度评分 user_love_score
- 规则如下：

```
1. If (rate is high) and (rank is very_high) then (user_love_score is good) (1)
2. If (rate is high) and (rank is high) then (user_love_score is good) (1)
3. If (rate is high) and (rank is middle) then (user_love_score is soso) (1)
4. If (rate is high) and (rank is low) then (user_love_score is soso) (1)
5. If (rate is high) and (rank is very_low) then (user_love_score is relatively_bad) (1)
```



```
6. If (rate is mid) and (rank is very__high) then (user_love_score is relatively_good) (1)
7. If (rate is mid) and (rank is high) then (user_love_score is relatively_good) (1)
8. If (rate is mid) and (rank is middle) then (user_love_score is soso) (1)
9. If (rate is mid) and (rank is low) then (user_love_score is relatively_bad) (1)
10. If (rate is mid) and (rank is very_low) then (user_love_score is relatively_bad) (1)
11. If (rate is low) and (rank is very__high) then (user_love_score is relatively_good) (1)
12. If (rate is low) and (rank is high) then (user_love_score is soso) (1)
13. If (rate is low) and (rank is middle) then (user_love_score is soso) (1)
14. If (rate is low) and (rank is low) then (user_love_score is bad) (1)
15. If (rate is low) and (rank is very_low) then (user_love_score is bad) (1)
```

Rule Base 3

- 用途：根据年代 year 和用户对年代的偏好 year_prefer 计算年代评分 year_score
- 规则如下：

```
1. If (year is 久远) and (user_prefer is 偏好久远) then (year_score is 高) (1)
2. If (year is 较久远) and (user_prefer is 偏好久远) then (year_score is 较高) (1)
3. If (year is 中间) and (user_prefer is 偏好久远) then (year_score is 一般) (1)
4. If (year is 较新) and (user_prefer is 偏好久远) then (year_score is 较低) (1)
5. If (year is 很新) and (user_prefer is 偏好久远) then (year_score is 低) (1)
6. If (year is 很新) and (user_prefer is 偏好新款) then (year_score is 高) (1)
7. If (year is 较新) and (user_prefer is 偏好新款) then (year_score is 较高) (1)
8. If (year is 中间) and (user_prefer is 偏好新款) then (year_score is 一般) (1)
9. If (year is 较久远) and (user_prefer is 偏好新款) then (year_score is 较低) (1)
10. If (year is 久远) and (user_prefer is 偏好新款) then (year_score is 低) (1)
11. If (year is 中间) and (user_prefer is 偏好中间) then (year_score is 高) (1)
12. If (year is 较久远) and (user_prefer is 偏好中间) then (year_score is 较高) (1)
13. If (year is 较新) and (user_prefer is 偏好中间) then (year_score is 较高) (1)
14. If (year is 久远) and (user_prefer is 偏好中间) then (year_score is 一般) (1)
15. If (year is 很新) and (user_prefer is 偏好中间) then (year_score is 一般) (1)
```

Rule Base 4

- 用途：根据清新、古典、浓郁、芬芳四个词语的喜好程度计算11种香调的推荐度
- 规则如下：

```
1. If (清新 is 喜欢) then (柑橘调 is high)(绿叶调 is high)(水生调 is high)(馥奇香调 is mid)(皮革调 is mid)(西普调 is mid)(木制调 is low)(东方调 is low)(美食调 is low)(花香调 is mid)(果香调 is mid) (1)
2. If (古典 is 喜欢) then (柑橘调 is mid)(绿叶调 is mid)(水生调 is mid)(馥奇香调 is high)(皮革调 is high)(西普调 is high)(木制调 is mid)(东方调 is mid)(美食调 is mid)(花香调 is low)(果香调 is low) (1)
3. If (浓郁 is 喜欢) then (柑橘调 is low)(绿叶调 is low)(水生调 is low)(馥奇香调 is
```

```
mid)(皮革调 is mid)(西普调 is mid)(木制调 is high)(东方调 is high)(美食调 is high)(花香调 is mid)(果香调 is mid) (1)
4. If (芬芳 is 喜欢) then (柑橘调 is mid)(绿叶调 is mid)(水生调 is mid)(馥奇香调 is low)(皮革调 is low)(西普调 is low)(木制调 is mid)(东方调 is mid)(美食调 is mid)(花香调 is high)(果香调 is high) (1)
5. If (清新 is 无感) then (柑橘调 is mid)(绿叶调 is mid)(水生调 is mid) (0.5)
6. If (古典 is 无感) then (馥奇香调 is mid)(皮革调 is mid)(西普调 is mid) (0.5)
7. If (浓郁 is 无感) then (木制调 is mid)(东方调 is mid)(美食调 is mid) (0.5)
8. If (芬芳 is 无感) then (花香调 is mid)(果香调 is mid) (0.5)
9. If (清新 is 讨厌) then (柑橘调 is low)(绿叶调 is low)(水生调 is low)(馥奇香调 is mid)(皮革调 is mid)(西普调 is mid)(木制调 is high)(东方调 is high)(美食调 is high)(花香调 is mid)(果香调 is mid) (0.8)
10. If (古典 is 讨厌) then (柑橘调 is mid)(绿叶调 is mid)(水生调 is mid)(馥奇香调 is low)(皮革调 is low)(西普调 is low)(木制调 is mid)(东方调 is mid)(美食调 is mid)(花香调 is high)(果香调 is high) (0.8)
11. If (浓郁 is 讨厌) then (柑橘调 is high)(绿叶调 is high)(水生调 is high)(馥奇香调 is mid)(皮革调 is mid)(西普调 is mid)(木制调 is low)(东方调 is low)(美食调 is low)(花香调 is mid)(果香调 is mid) (0.8)
12. If (芬芳 is 讨厌) then (柑橘调 is mid)(绿叶调 is mid)(水生调 is mid)(馥奇香调 is high)(皮革调 is high)(西普调 is high)(木制调 is mid)(东方调 is mid)(美食调 is mid)(花香调 is low)(果香调 is low) (0.8)
```

(四) 对模糊集、模糊规则和过程进行编码

1.编码工具为Fuzzy Logic Toolbox，编码成功后存储在fis类型的文件里 2.Rule Base 1 和 2 可直接在 MATLAB 工作台调用，批量进行输入输出 3.由于 Rule Base 3 和 4 需要用到用户在前端填写的数据，故需将其包装成 matlab 函数，以方便后端调用及数据传输 4.使用模糊推理引擎一共得到了14个评分值，需要在后端根据权重计算出最终的推荐度 5.最终推荐度 = $0.1 * s1 + 0.2 * s2 + 0.1 * year_score + 0.6 * style_score$

- $s1 = weight1 * user_love_score$ (weight1 为用户在前端填写的“用户喜爱度偏好”)
- $s2 = weight2 * time_score$ (weight2 为用户在前端填写的“留香时间偏好”)
- style_score 是根据每款香水的香调组成以及 11 种香调的推荐度综合得出的分数，具体逻辑详见代码

(五) 评估并调试系统

1. 重新定义输入输出变量的范围
2. 更改模糊集的形状和指标
3. 更改不同规则的权重值
4. 添加/删除/修改已有的规则
5.

六、总结

1. 学习了爬虫的相关知识，对数据挖掘和数据处理有了新的认识
2. 学习了 Matlab 的使用，特别是 Fuzzy Logic Toolbox 这一工具，对模糊专家系统的组成、运行逻辑、设计、调试等掌握得更为透彻