**CBR+PGPE推理人机交互界面**

**一、总体使用顺序：**

1、首先使用“2D mountainCar控制界面”训练二维环境中的训练策略，并显示性能曲线。

2、再在“3D mountainCar控制界面”中训练三维环境中的训练策略，并显示性能曲线。

3、最后点击“结果对比”按钮观察我们的方法与传统DQN方法得到的结果的对比。

**二、2D mountainCar 控制界面的使用**

1、点击“现有规则”按钮，查看根据人的先验知识构建的规则，目前针对任务构建了6条规则。

2、根据规则，依次添加参数的均值和方差。其中“均值一”和“方差一”对应参数theta；“均值二”和“方差二”对应参数beta。

3、点击“添加参数”按钮，如果提示框显示“参数添加完成”，表明参数添加成功，否则需要根据提示修改输入的参数。在参数的选择上，theta的均值在0.01左右，方差在0.002左右；beta的均值在-0.7左右，方差在0.05左右是比较合理的。可以改变取不同的值查看相应效果。

4、点击“训练”按钮，如果未有异常提示，则等待30秒左右训练完成，训练完成后会显示相应的参数值和奖励值的变化曲线。

5、点击“添加规则”按钮，会在规则框中添加相应的规则，但是目前并未与训练程序相连接。

**三、3D mountainCar控制界面的使用**

1、点击“规则映射”按钮，查看根据人的先验知识构建的二维到三维环境的动作映射规则，初始时有1条规则。

2、添加参数的均值和方差。其中“均值一”和“方差一”对应参数theta，由于只有一个参数，只添加这两个文本框即可。在取值方面，theta的均值取0.4左右，方差取0.1左右。可以改变取不同的值查看相应效果。

3、点击“添加参数”按钮，如果提示框显示“参数添加完成”，表明参数添加成功，否则需要根据提示修改输入的参数。

4、点击“训练”按钮，如果未有异常提示，此时系统会根据基于案例推理的方法进行训练，等待30秒左右，训练完成后会显示相应的参数值和奖励值的变化曲线。

5、点击“生成规则”按钮，系统会在初始点进行探索，等待20秒左右，系统会自动生成一条新规则显示在规则框中，并在“均值二”和“方差二”文本框中显示对应参数beta的均值和方差。这时**不需要**人为手动输入beta的均值和方差。

6、点击“训练”按钮，等待30秒左右，训练完成后会显示相应的参数值和奖励值的变化曲线。可以查看此时添加新的规则后的奖励值变化情况。

7、此时提示框中会显示“新规则整合完成”，规则框中会显示整合后的规则，共11条规则，是在不使用案例库的情况下的动作策略。

8、再次点击“训练”按钮，等待3~5分钟左右训练完成，训练完成后会显示相应的参数值和奖励值的变化曲线。

9、点击“结果对比”按钮，观察我们的方法与传统DQN方法得到的结果的对比。

10、点击“添加规则”按钮，会在规则框中添加相应的规则，但是目前并未与训练程序相连接。