



# 模仿你的行为

利节



## AIDemo示例

<http://aibook.csit.org>

# 实验准备：

**AIDemo** 提供了一个基于强化学习的 **Doom** 游戏示例程序 **fps-game**，以帮助读者理解强化学习的强大能力。这一示例程序是一个会打 **Doom** 游戏的 **AI** 机器人，基于深度强化学习方法训练得到。



# 实验原理：

训练时利用**MC** 采样生成训练样本，训练的目标是基于当前状态预测某一动作的未来收益。这一模型称为**DFP** (**Direct Future Prediction**) 模型。和很多深度强化学习方法不同的是，**DFP** 引入了一个任务向量 (**Goal Vector**)，基于该向量可灵活定义动作的收益。在我们的示例程序中，这一任务向量包括三个元素：弹药、生命和杀怪数，任务向量即是这三个元素在动作收益中的比重。在训练时，选择各种可能的任务向量使得模型适应不同的收益偏好；在运行时，定义不同的任务向量，即可以改变机器人的行为方式。



# 实验一：运行缺省配置

进入 `fps-game` 目录，运行 `run.sh`，即可启动主程序。这一程序导入一个先期训练完成的模型，在 `vizDoom` 模拟器中运行该模型。



## 实验二：改变任务向量

修改`code/examples/run_exp.py` 文件中的任务向量，观察改变目标后机器人的动作方式。例如：

```
experiment_args['test_objective_coeffs_meas'] =  
np.array([0.5, 0.5, 1.0])
```

上述向量中的三个数值分别代表三个目标：弹药、生命和杀怪数，设定值在-1 到1 之间。设定值越大，对应的目标将被实现的越好。比如，当设定为`[1.0, 1.0, -1.0]` 时，机器人将会更加珍惜子弹，维持生命，但尽量不击杀怪物。





The end !