有关强化基本原理的技术文章学习（它是什么以及它如何工作）及其在在线机器学习中的应用入侵检测的方法。 您可以使用一个或多个示例

需要描述的概念包括agent、状态、奖励、学习过程以及如何根据马尔科夫决策过程建模客观环境。

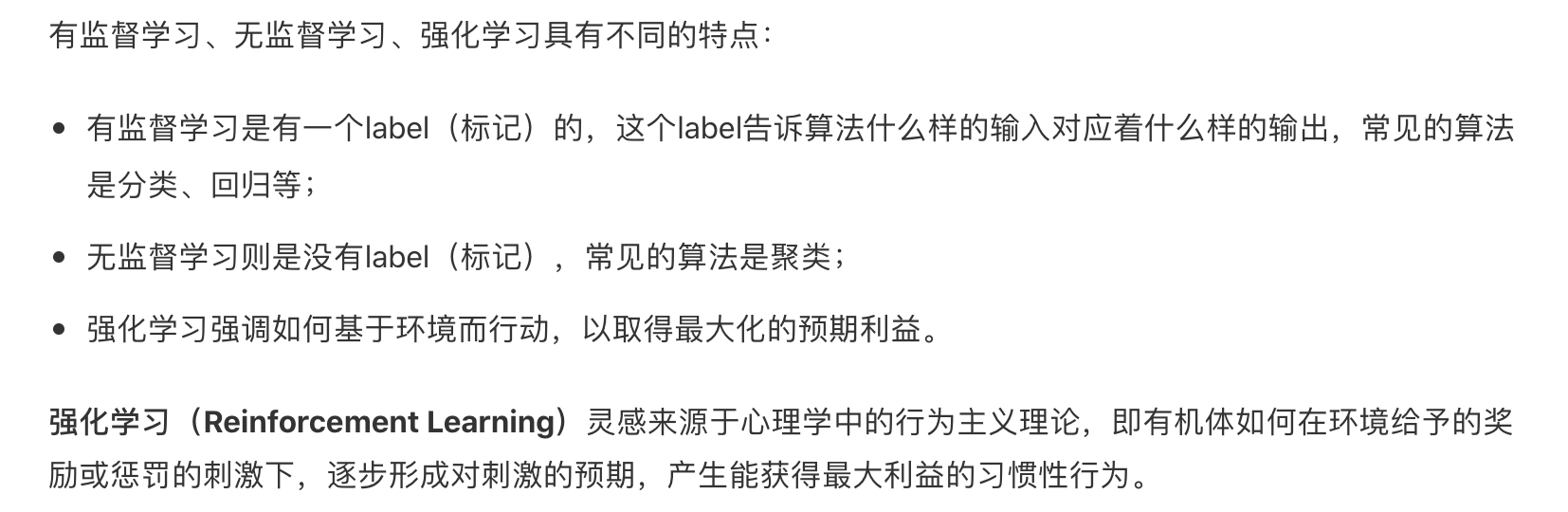
Agent: Agent is a putative entity, usually as an object trained in a particular environment to get the right decision. As an example, agent is a robot moving through a maze without any collisions.

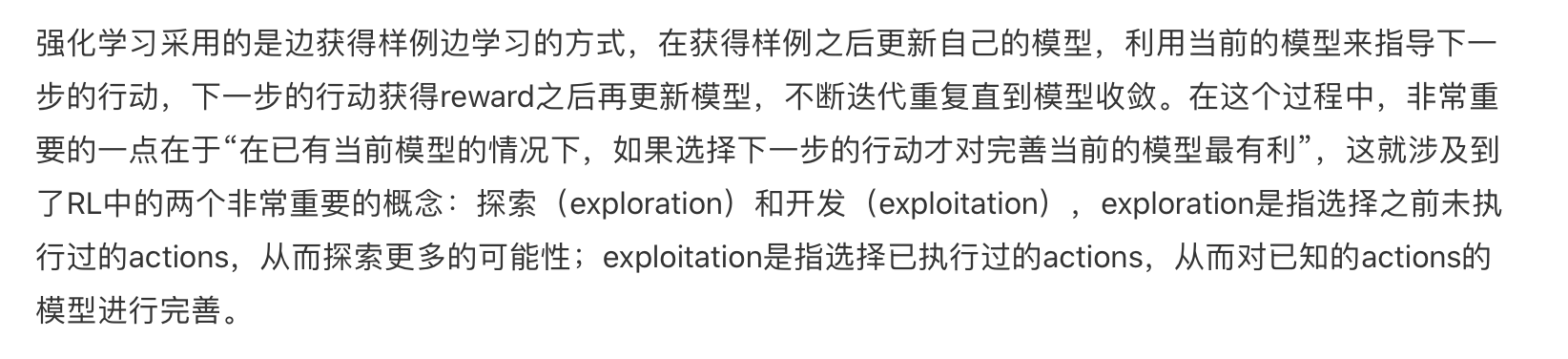
State: State defines the current real-time situation information of the Agent, such as the position of the robot in the maze, its movement posture or body twist Angle, etc. The state information of the Agent completely depends on the way to solve the problem

Reward: When the Agent receives real-time feedback after performing a specific behavior action or task, reward will be considered as a scalar, which can be divided into good and bad ones for performing the behavior in the current environment. For example, the score decreases when the robot touches a wall.

<https://blog.csdn.net/hellocsz/article/details/80835542>

<https://zhuanlan.zhihu.com/p/25319023>





强化学习的目的是最大化长期未来奖励，即寻找最大的 U。（注：回报也作 G 表示）

价值函数用来衡量某一状态或动作-状态的优劣，即对智能体来说是否值得选择某一状态或在某一状态下执行某一动作

我们需要找到最优的策略使未来回报最大化，求解过程大致可分为两步,具体内容会在后面展开

1. 预测：给定策略，评估相应的状态价值函数和状态-动作价值函数
2. 行动：根据价值函数得到当前状态对应的最优动作

Bellman 期望方程