随笔-16 文章-1 评论-0

增强学习贪心算法与Softmax算法

（一）$\varepsilon-$贪心算法

这个算法是基于一个概率来对探索和利用进行折中：每次尝试时，以ε概率进行探索，即以均匀概率随机选取一个摇臂，以$1-\varepsilon$的概率进行利用，即以这个概率选择当前平均奖赏最高的摇臂（如有多个，则随机选取）。

论文
前端

随笔档案(16)
2016年7月 (6)
2016年6月 (6)
2016年5月 (4)

文章分类(1)
Python
机器学习(1)
其他

阅读排行榜
1. 增强学习----介绍(1661)
2. 增强学习贪心算法与Softmax算法(910)
3. 增强学习————K-摇臂赌博机(579)
4. 对MVC的理解(165)
5. python画图(140)

令 $Q(k)$ 记录摇臂 k 的平均奖赏. 若摇臂 k 被尝试了 n 次, 得到的奖赏为 $v_1, v_2, \ldots, v_n$, 则平均奖赏为

$$Q(k) = \frac{1}{n}\sum_{i=1}^{n} v_i \ . \tag{16.1}$$

若直接根据式(16.1)计算平均奖赏, 则需记录 n 个奖赏值. 显然, 更高效的做法是对均值进行增量式计算, 即每尝试一次就立即更新 $Q(k)$. 不妨用下标来表示尝试的次数, 初始时 $Q_0(k) = 0$. 对于任意的 $n \geqslant 1$, 若第 $n-1$ 次尝试后的平均奖赏为 $Q_{n-1}(k)$, 则在经过第 n 次尝试获得奖赏 v_n 后, 平均奖赏应更新为

$$Q_n(k) = \frac{1}{n}\big((n-1) \times Q_{n-1}(k) + v_n\big) \tag{16.2}$$

$$= Q_{n-1}(k) + \frac{1}{n}\big(v_n - Q_{n-1}(k)\big) \ . \tag{16.3}$$

其中：小k表示第k个摇臂。因为大K表示摇臂总数；n表示尝试的次数，vn表示第n次尝试的奖赏。

Qn的直观意思为：Q_{n-1}为前n-1次的平均奖赏。当其与n-1相乘所得是前n-1次总奖赏。再加上第n次的奖赏，处于n，则为n次的平均奖赏。

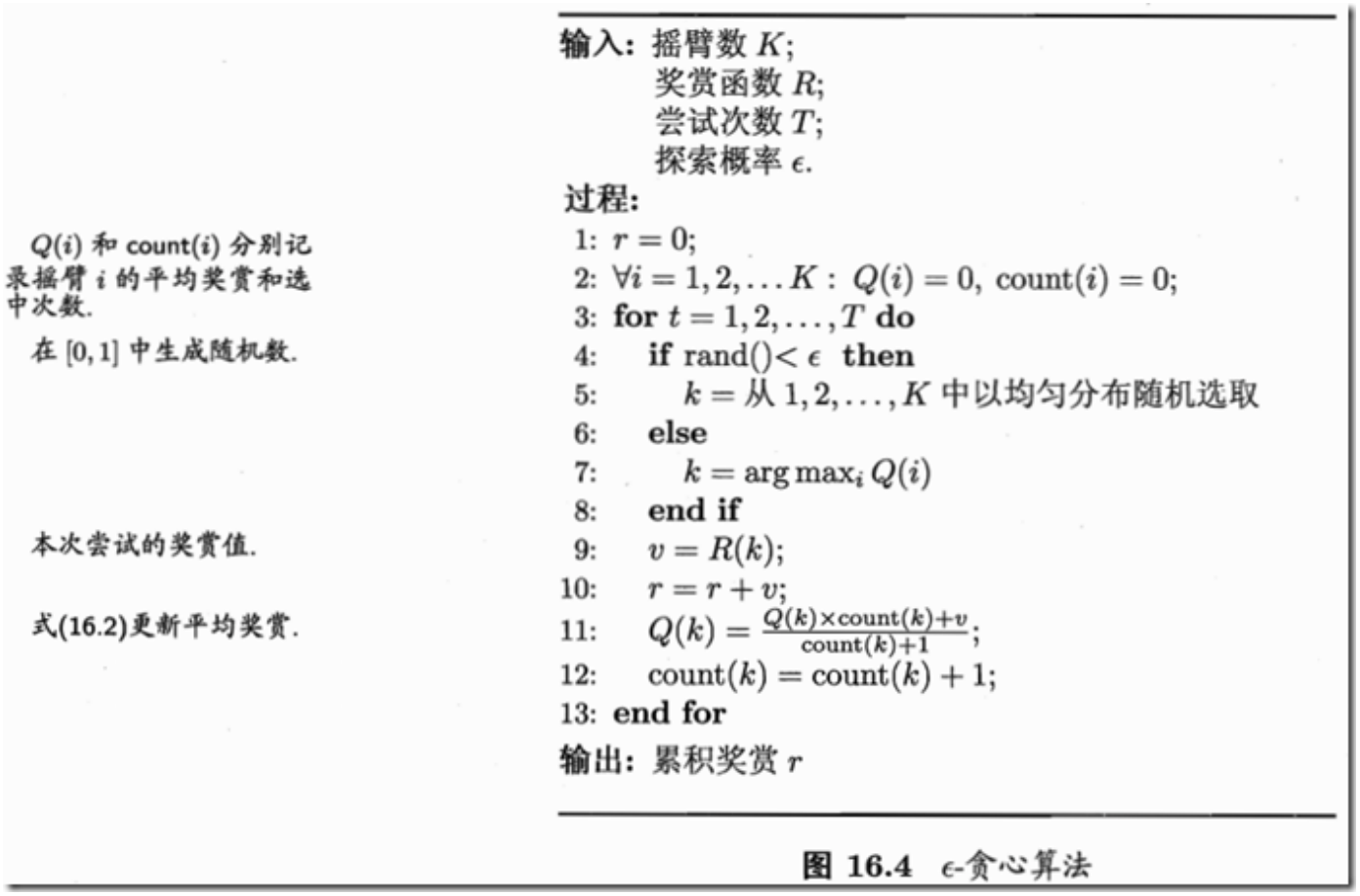


图 16.4 ϵ-贪心算法

其中：argmax为选取最优的Q(i)。count是从0开始的，故count(k)+1的值为n，算出的Q(k)为n次的平均奖赏。

（二）Softmax算法

softmax算法是基于当前已知的摇臂平均奖赏来对探索和利用进行折中。若各摇臂的平均奖赏相当，则选取各摇臂的概率也相当；如果某些概率的平均奖赏明显高于其他奖赏，则它们被选的概率也明显高。

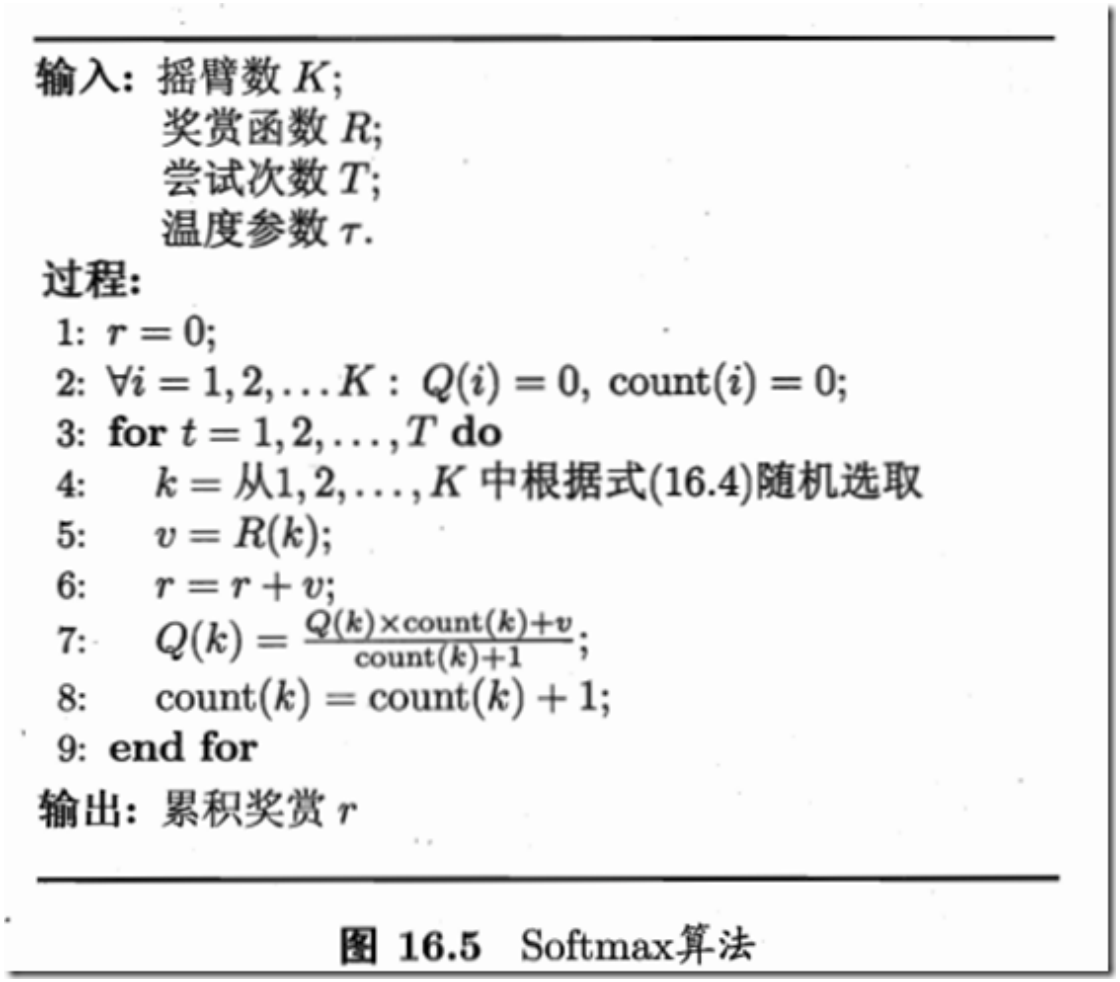
在贪心算法中，ε的取值是由用户选取的。softmax算法中摇臂概率的分配是基于Boltzmann分布

$$P(k)=\frac{e^{\frac{Q(k)}{\tau}}}{\sum_{i=1}^{K}e^{\frac{Q(i)}{\tau}}},$$

。

<查>Boltzmann分布

其中，$Q(i)$ 记录当前摇臂的平均奖赏；$\tau>0$ 称为“温度”，τ 越小则平均奖赏高的摇臂被选取的概率越高. τ 趋于 0 时 Softmax 将趋于“仅利用”，τ 趋于无穷大时 Softmax 则将趋于“仅探索”. Softmax 算法描述如图 16.5 所示.



输入：摇臂数 K;
奖赏函数 R;
尝试次数 T;
温度参数 τ.

过程：

1: $r = 0$;
2: $\forall i = 1, 2, \ldots K : Q(i) = 0, \text{count}(i) = 0$;
3: **for** $t = 1, 2, \ldots, T$ **do**
4: $k =$ 从$1, 2, \ldots, K$ 中根据式(16.4)随机选取
5: $v = R(k)$;
6: $r = r + v$;
7: $Q(k) = \frac{Q(k) \times \text{count}(k) + v}{\text{count}(k) + 1}$;
8: $\text{count}(k) = \text{count}(k) + 1$;
9: **end for**

输出：累积奖赏 r

图 16.5 Softmax算法

从算法中并没有看出botlzmann分布的利用？

至于两个算法的取舍还要看实际情况。从下图看出，softmax当温度=0.01时，曲线与“仅利用”的曲线几乎重合。

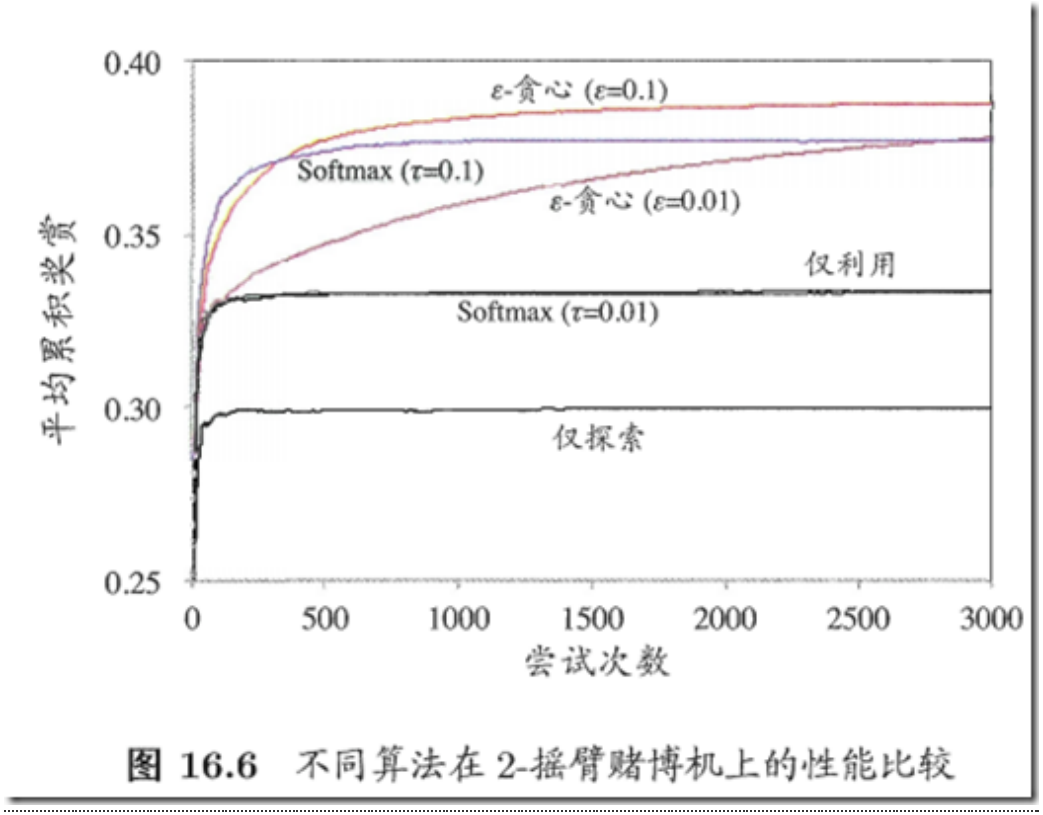


图 16.6 不同算法在 2-摇臂赌博机上的性能比较

分类: 机器学习

标签: 增强学习, 强化学习

好文要顶 关注我 收藏该文

YC_Yuan
关注 - 3
粉丝 - 0
+加关注

0 0

» 下一篇：增强学习————K-摇臂赌博机

posted @ 2016-05-31 09:12 YC_Yuan 阅读(909) 评论(0) 编辑 收藏