Page 4, Count-based methods

This could be implemented as an intrinsic reward which is added to the reward obtained from the environment (which is called extrinsic reward in this context). One of the options to formulate such a reward is to use the Bandits Exploration approach (which is an important variation of the RL problem): $r_i = c\frac{1}{\sqrt{\tilde{N}(s)}}$. Here, $\tilde{N}(s)$ is a count or pseudo-count of times weve seen the state s, and value c defines the weight of the intrinsic reward.

If the amount of states is small, like in the tabular learning case, we can just count them. In more difficult cases, when there are too many states, some transformation of the state needs to be introduced, like the hashing function or some embeddings of the states. Another approach to count the states is called pseudo-count methods, when $\tilde{N}(s)$ is factorized into the density function and the total amount of states visited.