基于K最近邻训练的深度Q学习

2018年3月13日

1 算法

Q学习的最优动作-价值函数：

$$Q^*(s,a) = \mathbf{E}_{s'\sim\xi}[r + \gamma \cdot max_{a'} Q^*(s^{'}, a^{'})|s,a] \tag{1}$$

$$L_i(\theta_i) = \mathbf{E}_{s,a\sim\rho(\cdot)}[(y_i - Q(s,a;\theta_i))^2] \tag{2}$$

$$\nabla_{\theta_i} L_i(\theta_i) = \mathbf{E}_{s,a\sim\rho(\cdot);s'\sim\xi}[(y_i - Q(s,a;\theta_i))\nabla_{\theta_i} Q(s,a;\theta_i)] \tag{3}$$

基于K最近邻训练深度Q学习方法中，设置一个K最近邻字典（KND，K-nearest Neighbourhood Dict），KND支持查询和写操作，通过状态key查询到K最近邻的激励集合V_a和状态集合K_a，状态key的激励$r = \sum_i w_i v_i$，v_i是V_i的第i个最近邻对应动作a的激励，$w_i = \frac{k(h,h_i)}{\sum_j k(h,h_j)}$，$h_i$是$K_a$ 的第i个最近邻对应的状态，$k(x,y)$是向量x和向量y的核函数，如高斯核函数$k(h,h_i) = \frac{1}{||h-h_i||_2^2} + \delta$。