算法概览图符号表

符号	英文含义	中文含义
t	timestep in real environment	在真实环
$\gamma \in [0,1]$	discount factor	折扣因子
k	hypothetical step	“假设步 结果在模
o_t	observation at timestep t	在时间步
u_t	reward at timestep t	在时间步
a_t	action at timestep t	在时间步
s_t	hidden state at timestep t	在时间步
r_t	predicted reward at timestep t	在时间步
p	predicted policy logits	预测的策
v	predicted value	预测的值
h	Representation Network	表征网络 hidden st
g	Dynamics Network	动力学网
f	Prediction Network	预测网络
vp	predicted value prefix	预测的值
r_h^k	Reward hidden state at hypothetical step k	在假设步 EfficientZ
g^{rnn}	(recurrent) Dynamics Network	（循环） $vp^k, s^k,$ EfficientZ
β	Proposal policy distribution	提议策略 用。
$\hat{\beta}$	empirical policy distribution	经验策略

$\tau = \{s_0, a_0, s_1, a_1, ...\}$	trajectory	
$R(\tau) = \sum_{t=1}^{T} r_t$	return	
$G_t = \sum_{k=0}^{T-t-1} \gamma^k r_{t+k}$	Discounted return	
$\pi_\theta(a_t \vert s_t)$	stochastic policy	随机
$V_\pi(s) = \mathbb{E}_\pi[G_t \vert s_t = s]$	state value fuction	
$Q_\pi(s, a) = \mathbb{E}_\pi[G_t \vert s_t = s, a_t = a]$	state-action value fuction	

(表1：算法概览图的符号说明。)