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
$\gamma \in [0,1]$				蒲源 6719	discount factor	折扣因子
k ⋒⋒6719	浦源 6719	蒲源 6719	蒲源 67.19	蒲源 6719	hypothetical step	"假设步
ot	蒲源 6719		蒲源 6719		observation at timestep t	在时间步
u_t	-10	蒲源 6719	m0	蒲源 6719	reward at timestep t	在时间步
a_t					action at timestep t	在时间步
st					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 119 and 6119	表征网络 hidden s
g	蒲源 6719	Av.	蒲源 6719	AV.	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 Efficient2
g^{rnn}		蒲源 6719	75 JR 01	蒲源 6719	(recurrent) Dynamics Network	(循环): vp^k,s^k,s^k
	浦源 6719		蒲源 6719		前源 6T19 前源 6T19	Efficient
β					Proposal policy distribution	提议策略
\hat{eta}	2M 192- 2	蒲源 6719	AUL NO.	蒲源 6719	empirical policy distribution	经验策略

蒲源 6719	$ au \ = \ \{s_0, a_0, s_1, a_1,\}$		trajectory				新展 6719	
浦源 6719	V 1	潮源 6719	return	痛源 6719	丽源 6719	清源 6719	浦原 6719	蒲
滅源 6719	$G_t = \sum_{k=0}^{T-t-1} \gamma^k r_{t+k}$		Discounted return				重原 6719	蒲
	$\pi heta \left(at \left st ight) ight.$		stochastic policy					随村
蒲源 6719	$V_{\pi}(s) \ = \ \mathbb{E}_{\pi}[G_t s_t=s]$		state value fuction				蒲原 6719	
	$Q_{oldsymbol{\pi}}(s,a) \ = \ \mathbb{E}_{oldsymbol{\pi}}[G_t s_t=s,a_t=$	a]	state-action value fu	ıction				

(表1: 算法概览图的符号说明。)