

% 学習した政策を利用するためのプログラム

```
function test(Q, options)
    nactions = 9; % 行動数
    step = 1; % ステップカウンタ

    while(1)
        % 状態, 報酬, ゲーム状況の観測
        [state, state3, fin]=observe_test;

        % 政策の生成
        policy = zeros(1,nactions);
        switch(options.pmode)
            case 1 % greedy
                [v,a] = max(Q(state,:));
                policy(a) = 1;

            case 2 % e-greedy
                [v,a] = max(Q(state,:));
                policy = ones(1,nactions)*options.epsilon/nactions;
                policy(a) = 1-options.epsilon+options.epsilon/nactions;

            case 3 % softmax
                policy=exp(Q(state,:)/options.tau)/sum(exp(Q(state,:)/options.tau));
            end

        % 行動の選択および実行
        [a] = action_test(policy, step, state3);
        step = step + 1;
    end
end
```