

# 強化学習を用いた ごみ収集プログラムの提案

星野研究室

1090262 近森公貴

# 目次

- 背景・目的
- 強化学習
- ごみ収集問題
  - 最短経路学習実験
  - ごみ削除実験
- 問題点・今後の課題
- まとめ

# 背景

- 近年、ロボットの活躍が期待される範囲が家庭などへ拡大している
- あらかじめ埋め込まれた機能だけではなく、ロボット自身が学習し、新しい機能を自ら身に付けていく能力が必要とされている

# 目的

- 本研究では**強化学習**を用いてごみを収集する最短経路を学習させる
- また、学習結果を用い、ある地点のごみを削除する場合に、効率良く学習できるための手法を提案し検証していく

# 強化学習

- 強化学習とは、報酬という特別な入力を用いて環境に適応する機械学習の一種
- 環境と相互作用しているエージェントを扱い、エージェント自らが試行錯誤を繰り返し学習を行う

# Q-learning

- Q-learningは機械学習手法の方策オフ型TD学習の一つである
- Q値は状態 $s$ における行動 $a$ の選択重みを表す
- 報酬を得た場合、その時間 $t$ における状態 $s_t$ と行動 $a_t$ のQ値を更新する

$$Q(s_t, a_t) \leftarrow (1 - \alpha)Q(s_t, a_t) + \alpha[r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})]$$

報酬  $r$  学習率  $0 < \alpha \leq 1$  割引率  $0 \leq \gamma \leq 1$  時間  $t$

# ボルツマン選択

- ボルツマン分布を用い、評価の高い行動には最も高い選択確率が与えられ、他の行動には相応の重みをかけて確率的に行動する手法
- 確率 $p(a|s)$ に従って行動をルーレット選択する

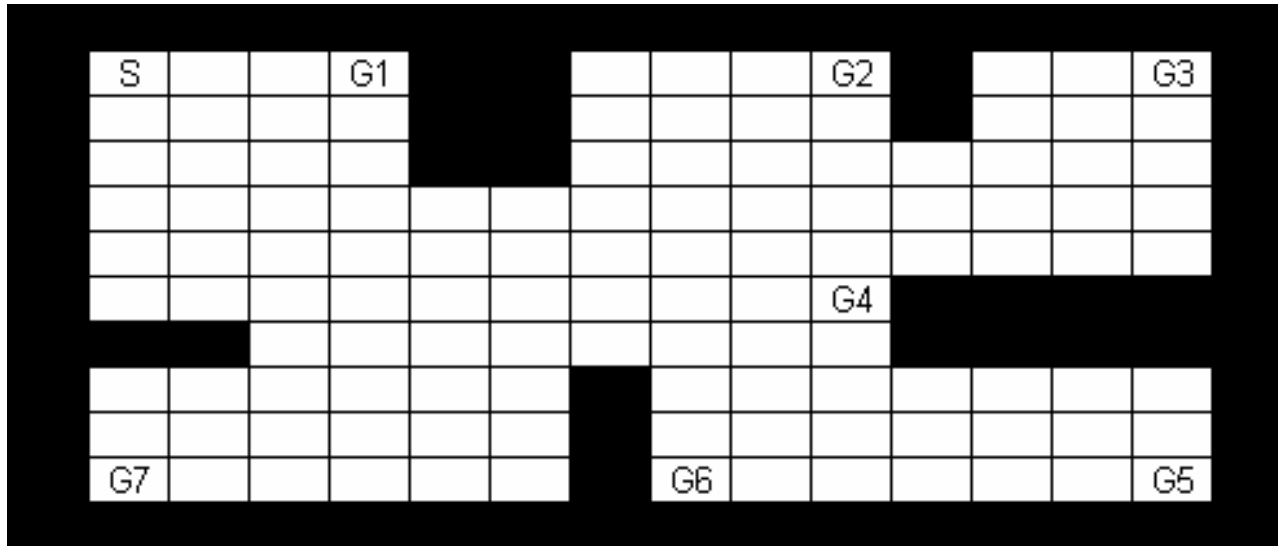
$$p(a | s) = \frac{\exp(\frac{Q(s, a)}{T})}{\sum_{b \in \text{action}} \exp(\frac{Q(s, b)}{T})}$$

温度定数  $T$

# ごみ収集問題

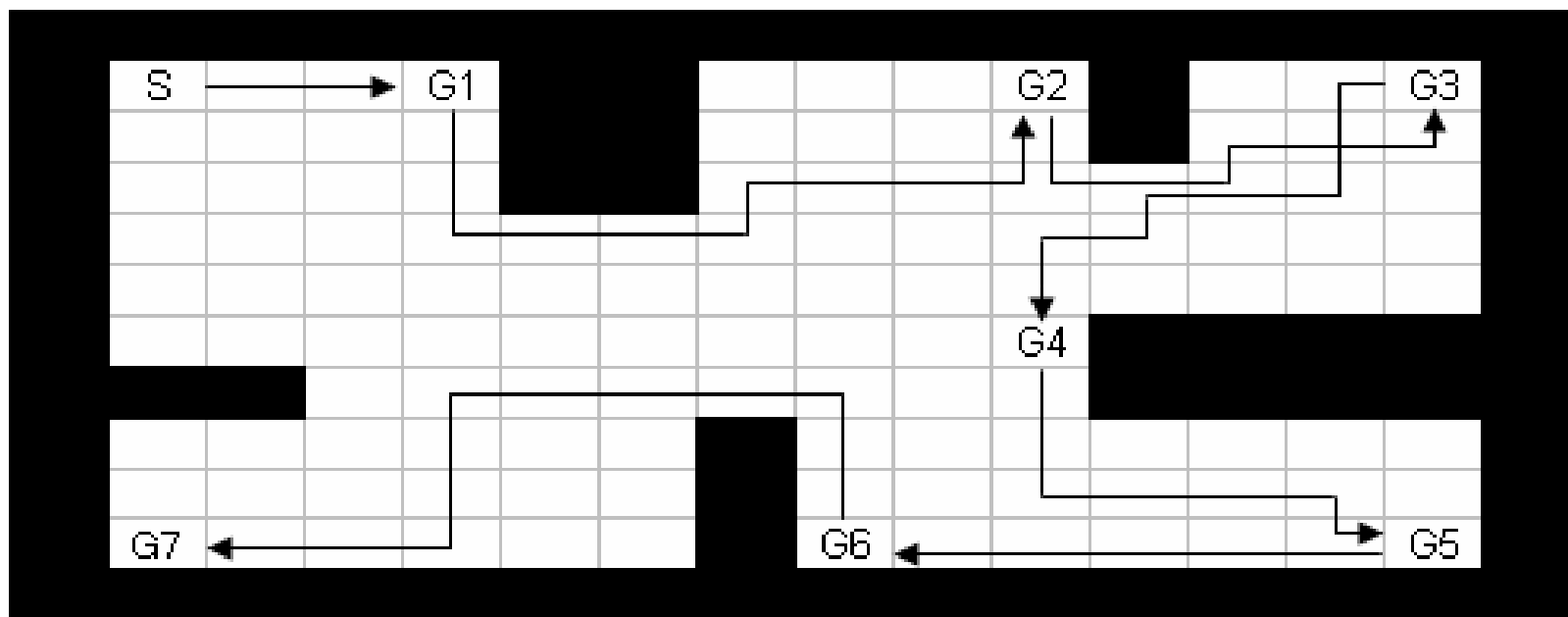
- ごみ収集問題とは、フィールドに置かれているごみをエージェントが回収していくというもの
- 強化学習を用いて経路を学習させる
  - Q-learningの学習率を0.1、割引率を0.9に設定
  - ごみを回収すると報酬10、壁に当たると-1を与える
  - エージェントは上下左右の4つの行動をとる
  - 行動選択にボルツマン選択を使用  
温度定数を0.06に設定

# 最短経路学習実験



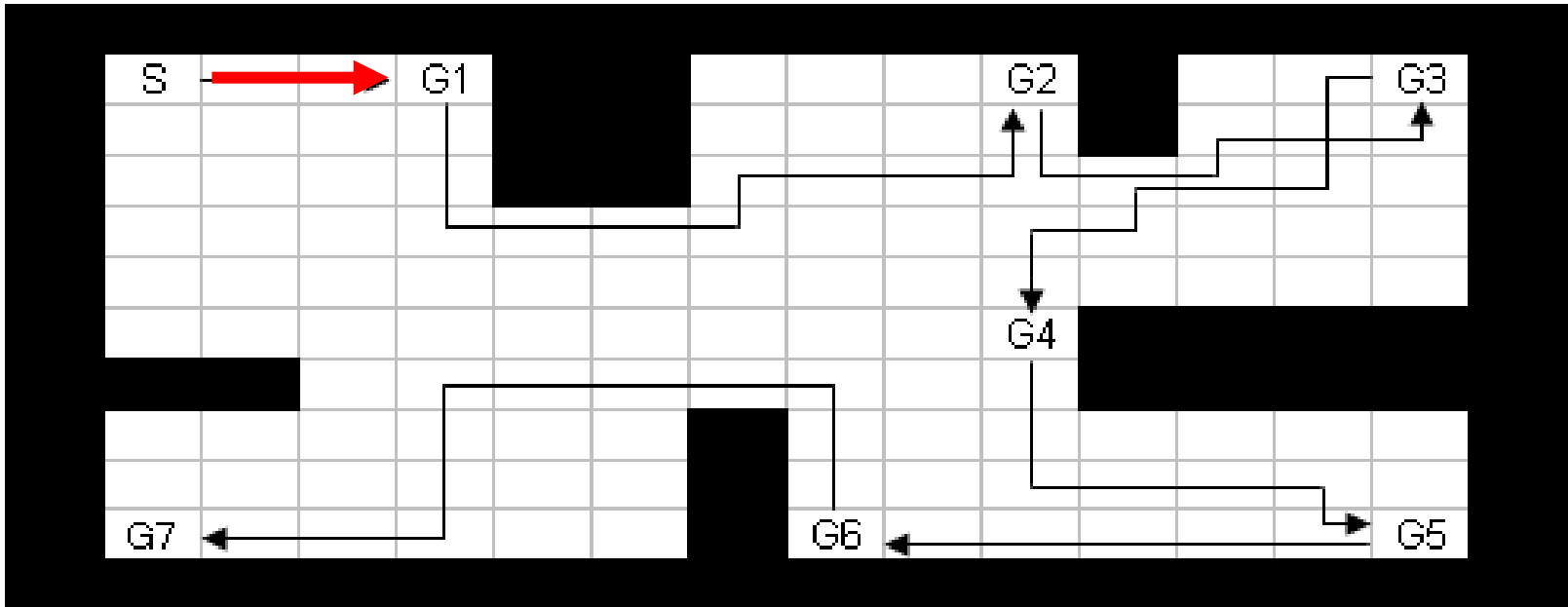
- ごみを収集した際の経路を学習
- Sがスタート、GがごみでG1～G7の順で回収する
- 学習はごみ毎に行う
- Q値はごみ毎に学習し保存した

# 最短経路学習実験結果



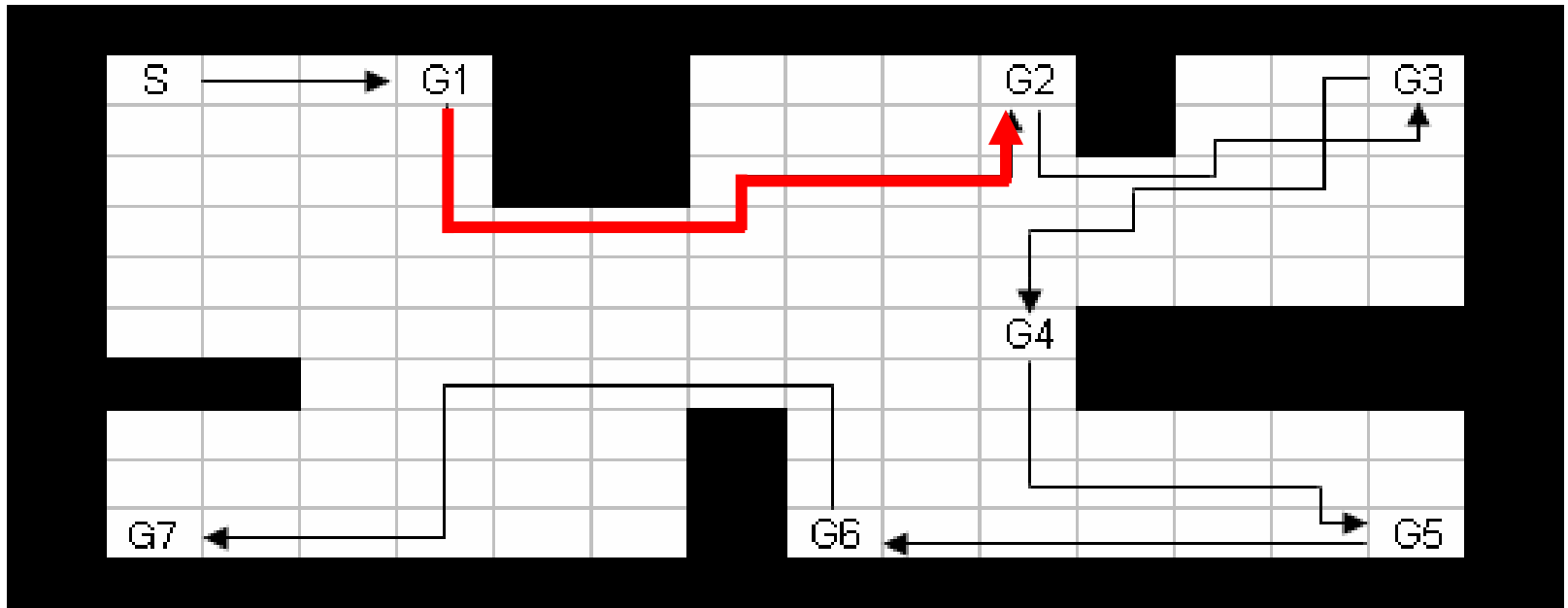
- 矢印の方向が学習した経路である
- 結果から、最短経路の1つを学習できた

# 最短経路学習実験結果



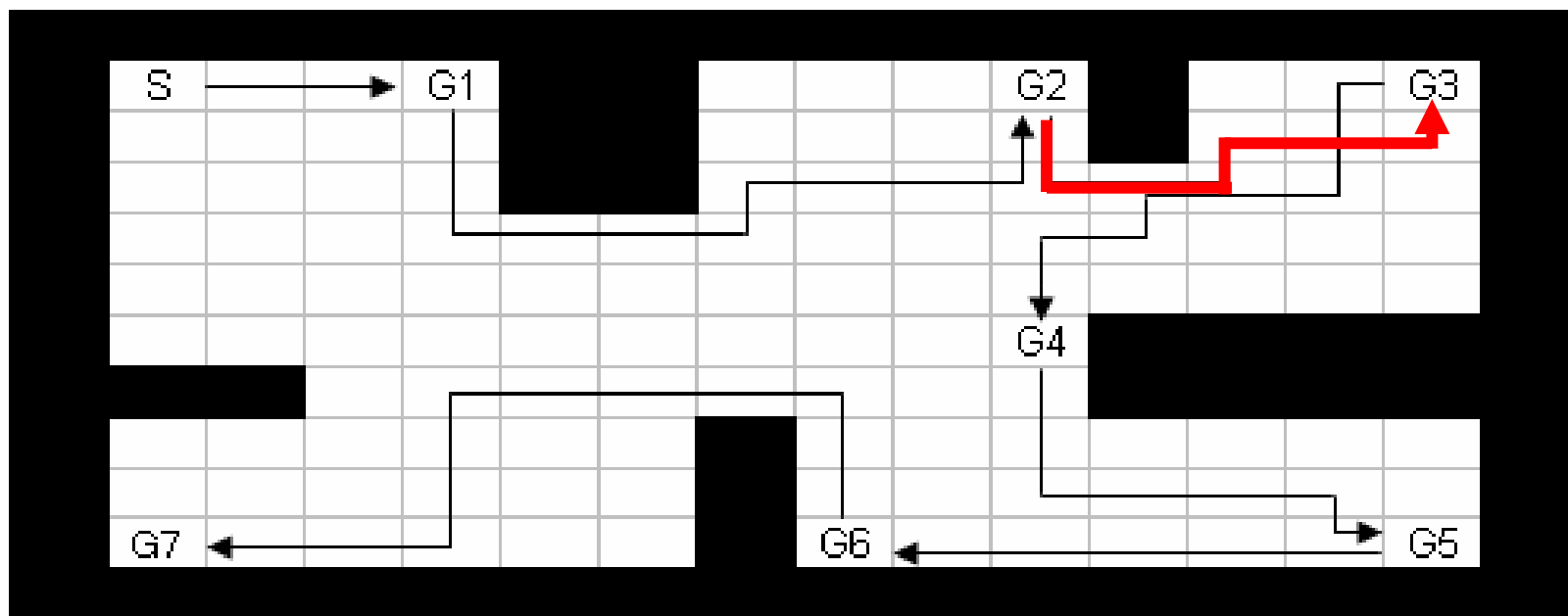
- 矢印の方向が学習した経路である
- 結果から、最短経路の1つを学習できた

# 最短経路学習実験結果



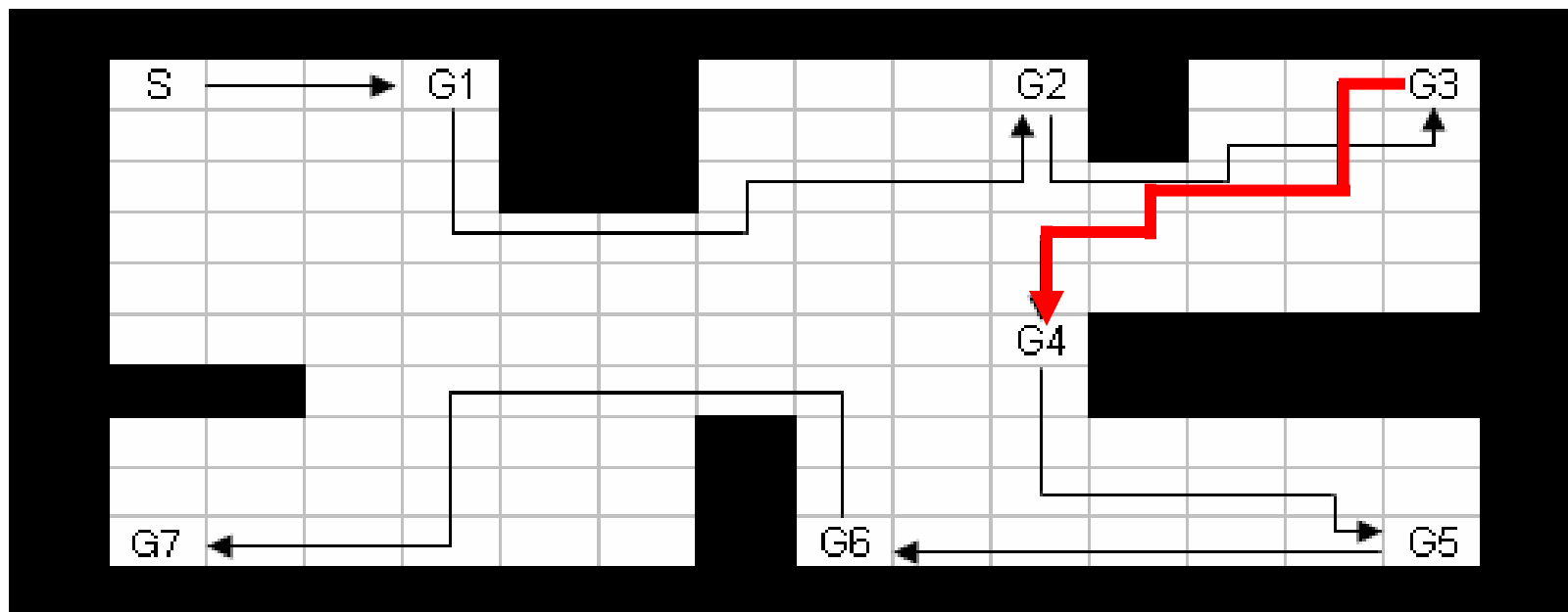
- 矢印の方向が学習した経路である
- 結果から、最短経路の1つを学習できた

# 最短経路学習実験結果



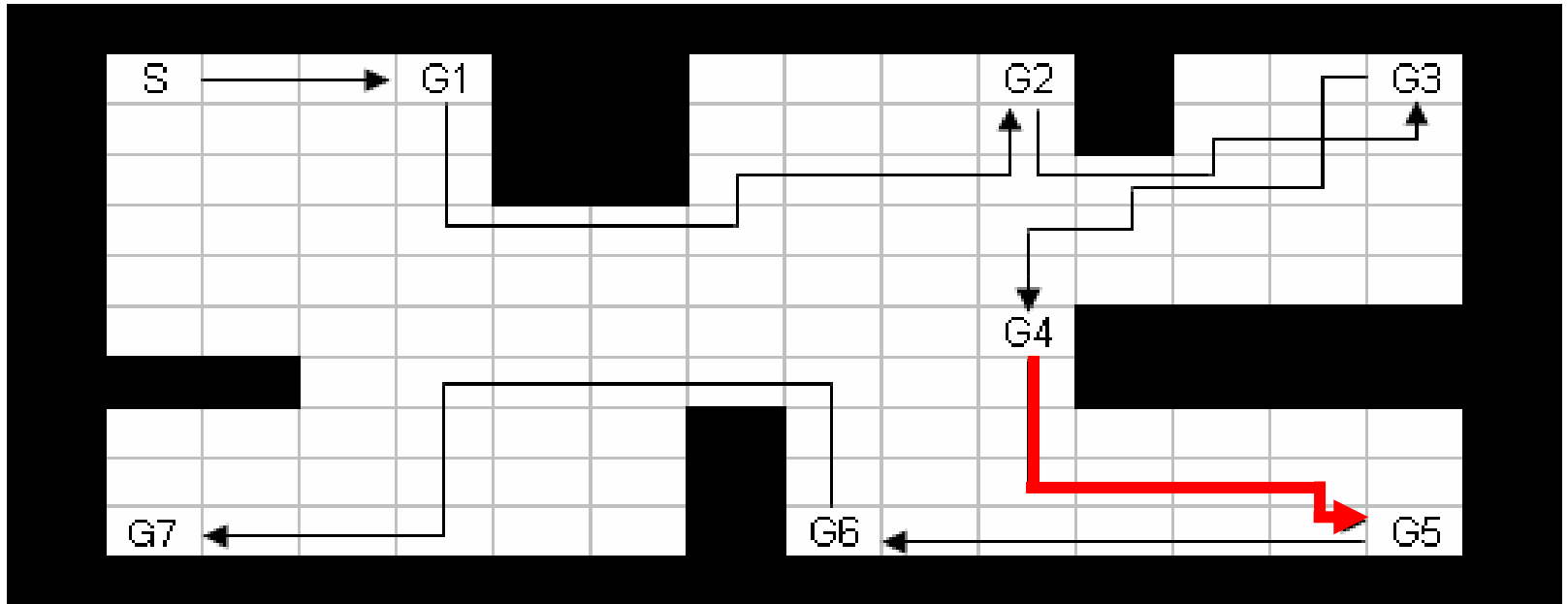
- 矢印の方向が学習した経路である
- 結果から、最短経路の1つを学習できた

# 最短経路学習実験結果



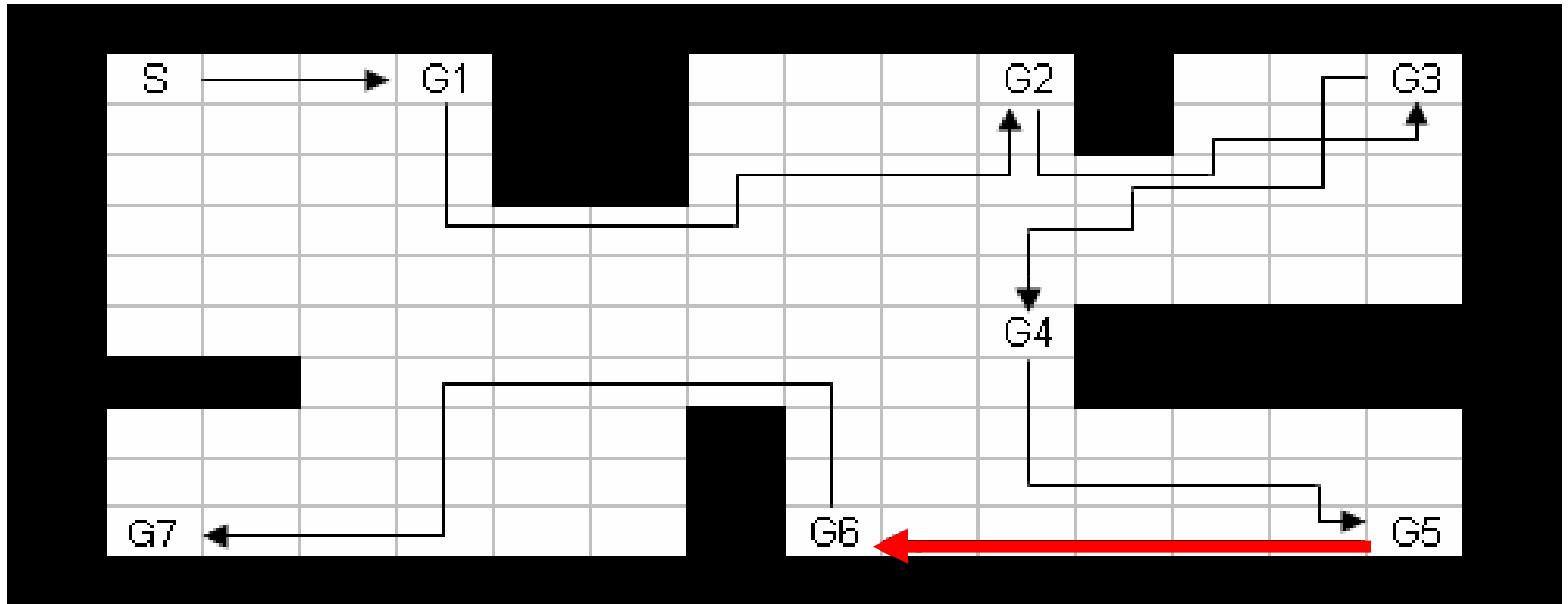
- 矢印の方向が学習した経路である
- 結果から、最短経路の1つを学習できた

# 最短経路学習実験結果



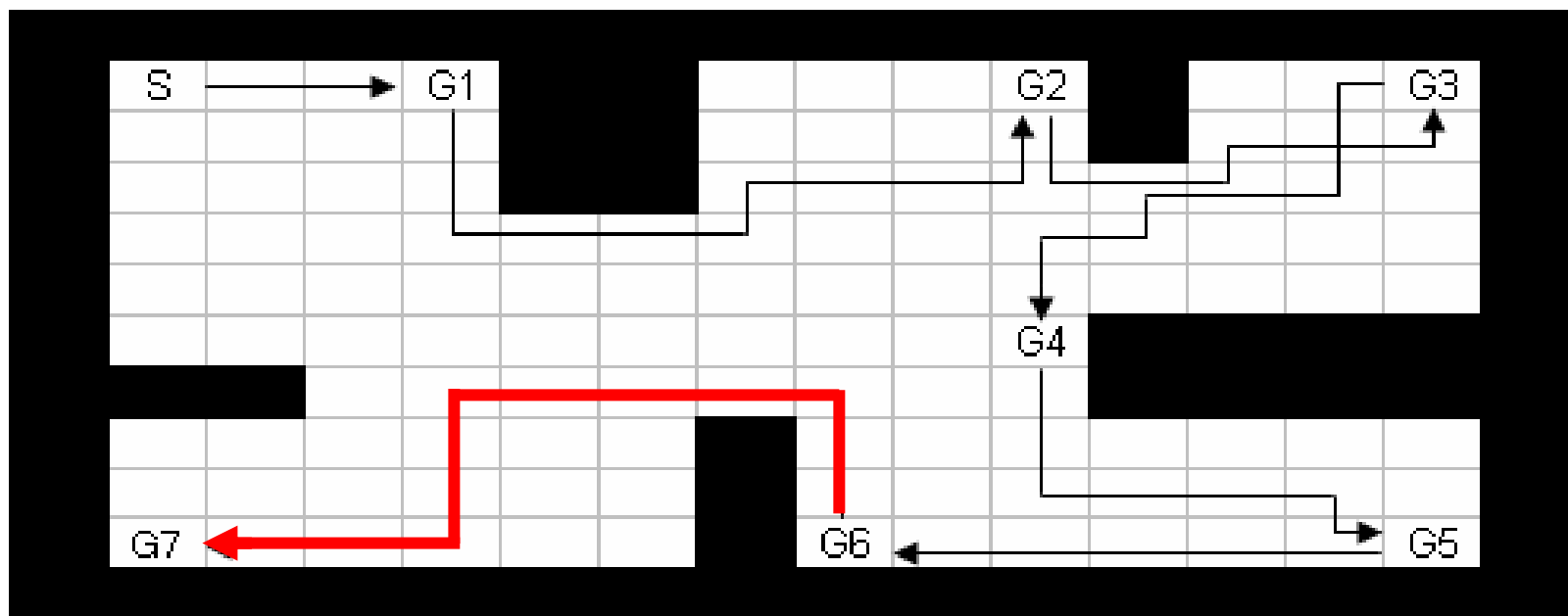
- 矢印の方向が学習した経路である
- 結果から、最短経路の1つを学習できた

# 最短経路学習実験結果



- 矢印の方向が学習した経路である
- 結果から、最短経路の1つを学習できた

# 最短経路学習実験結果

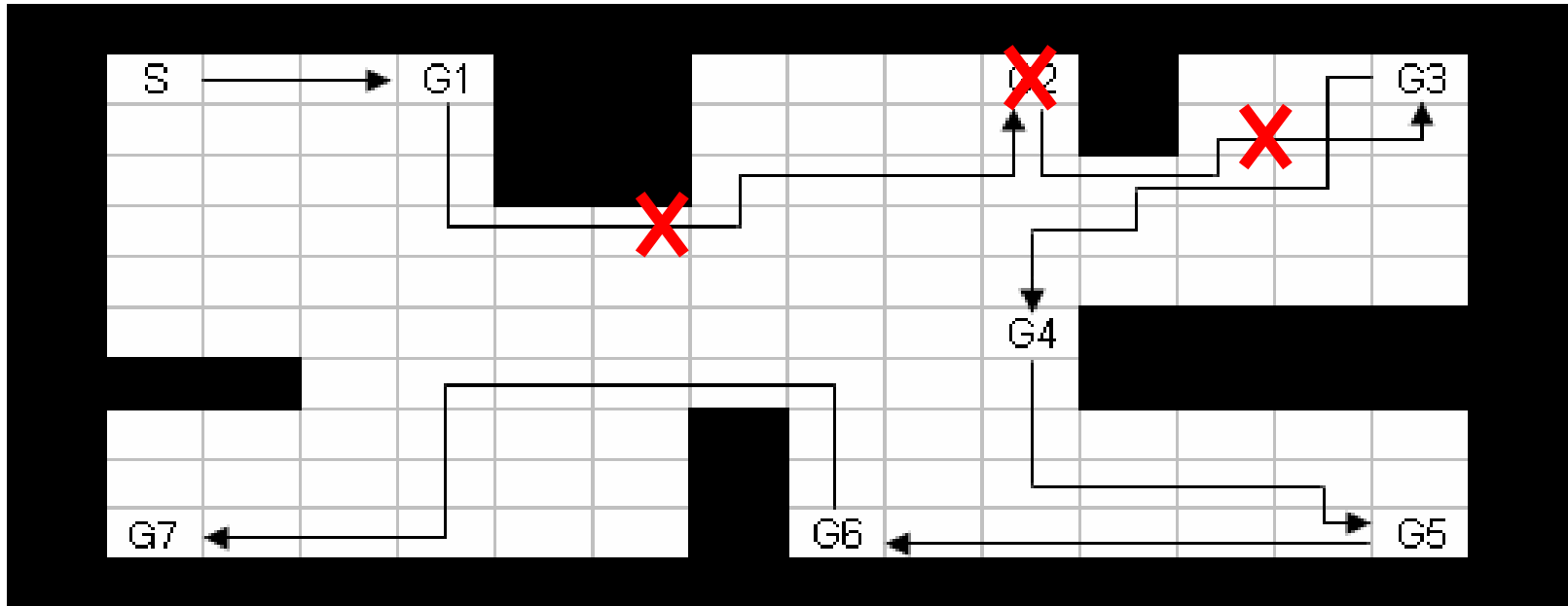


- 矢印の方向が学習した経路である
- 結果から、最短経路の1つを学習できた

# ごみ削除実験

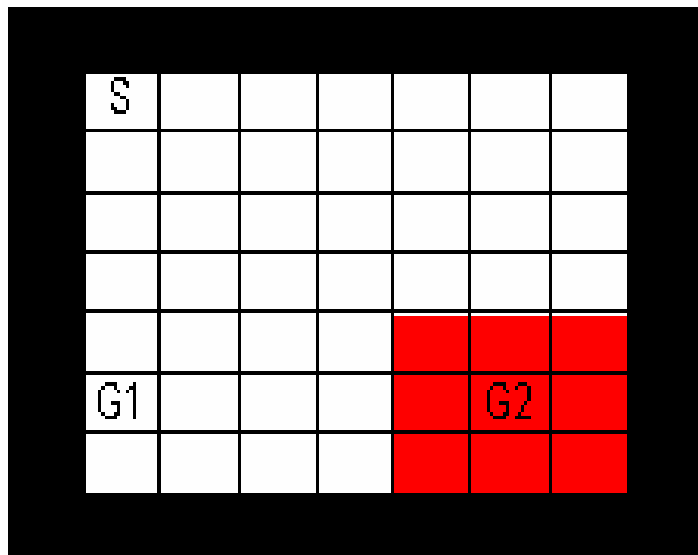
- $S \rightarrow G1, G1 \rightarrow G2 \dots$  のように各経路別にQ値を格納した
- 最短経路学習実験の結果から、ある地点のごみを削除した場合の**最短経路**を求める
- その際に学習効率を上げる方法を考え、提案手法とQ-learningの比較をする
- 今回の実験では、ごみG2を削除した場合の**経路**を学習させる

# ごみ削除実験結果



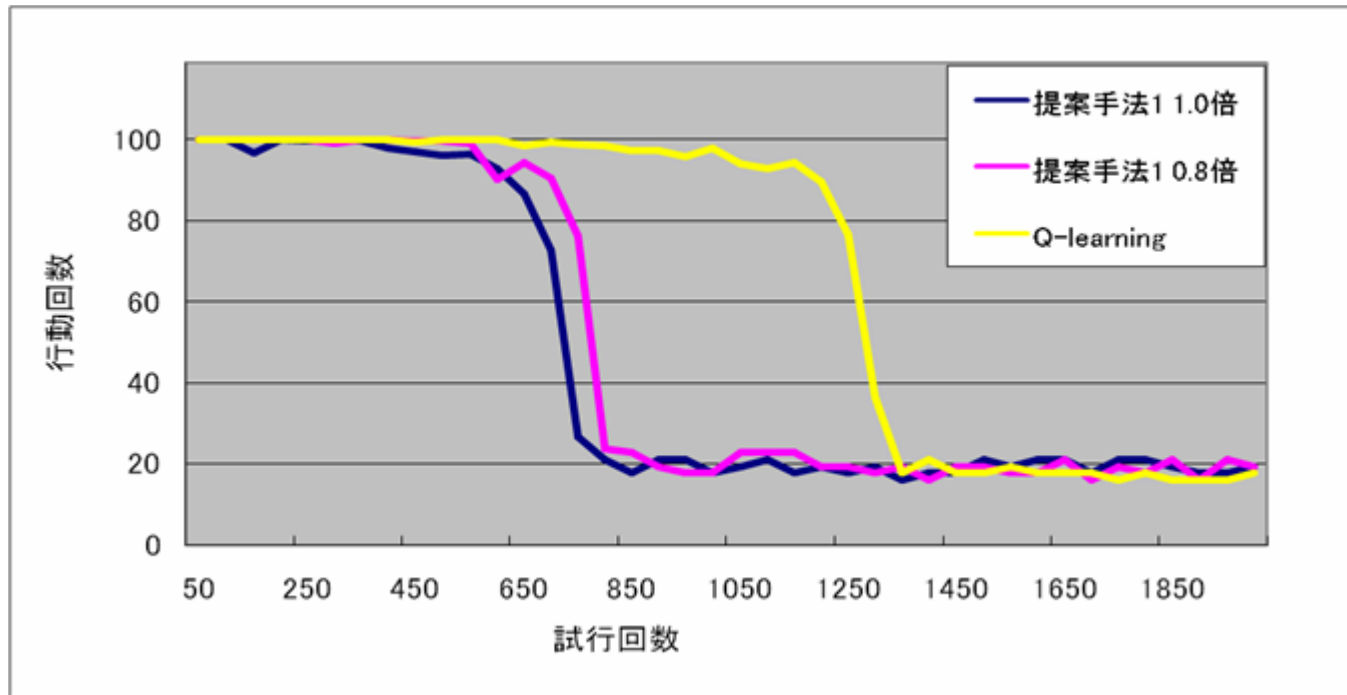
- **×** の部分の経路を再学習することになる
- G1→G3の**経路を学習させる**

# 提案手法1



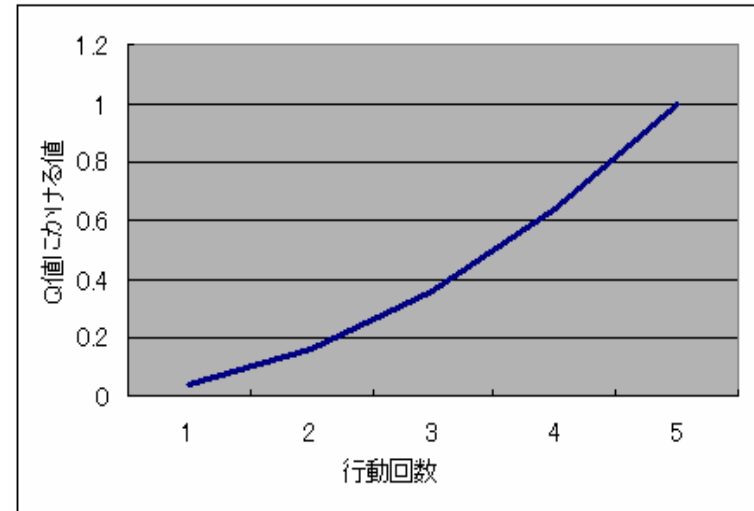
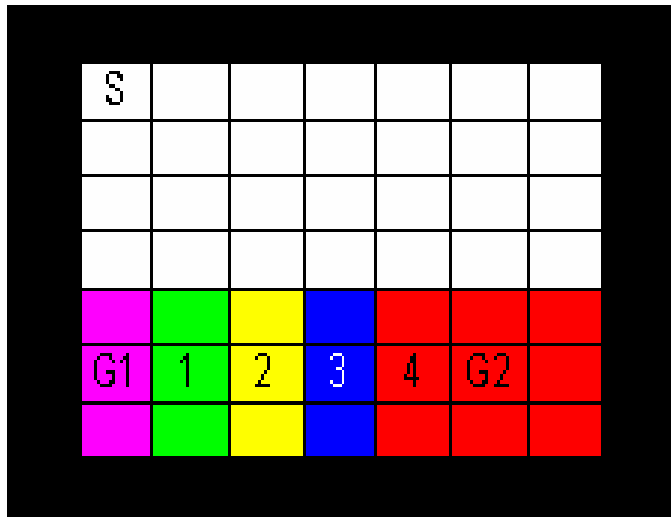
- G1を削除する場合、G1からG2までの学習したQ値をG2の周りだけ残す
- 残したQ値を0.8倍したものと1.0倍したものを初期値として、SからG2までの経路を学習させる

# 提案手法1の結果



- Q-learningに比べ提案手法1は少ない試行回数で学習できた
- 提案手法0.8、1.0倍ともに同様の結果となった

# 提案手法2



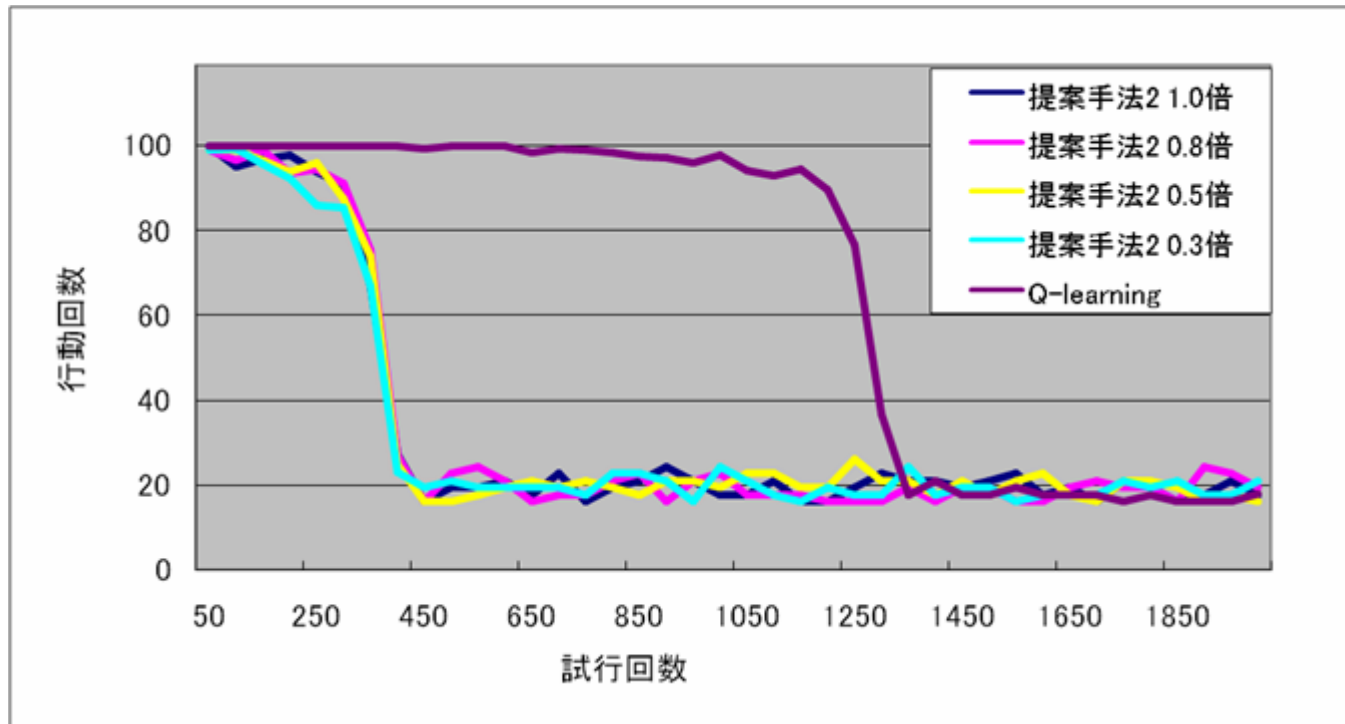
- G2の周りだけではなくG1→G2の学習した最短経路に対応するそれぞれのQ値を格納した
- 次に格納したQ値に $y$ の値をかけ、さらに0.3～1.0倍して割り引いた値を初期値とし、S→G2の経路を学習させた

$$y = \frac{x^2}{n^2}$$

行動回数  $x$

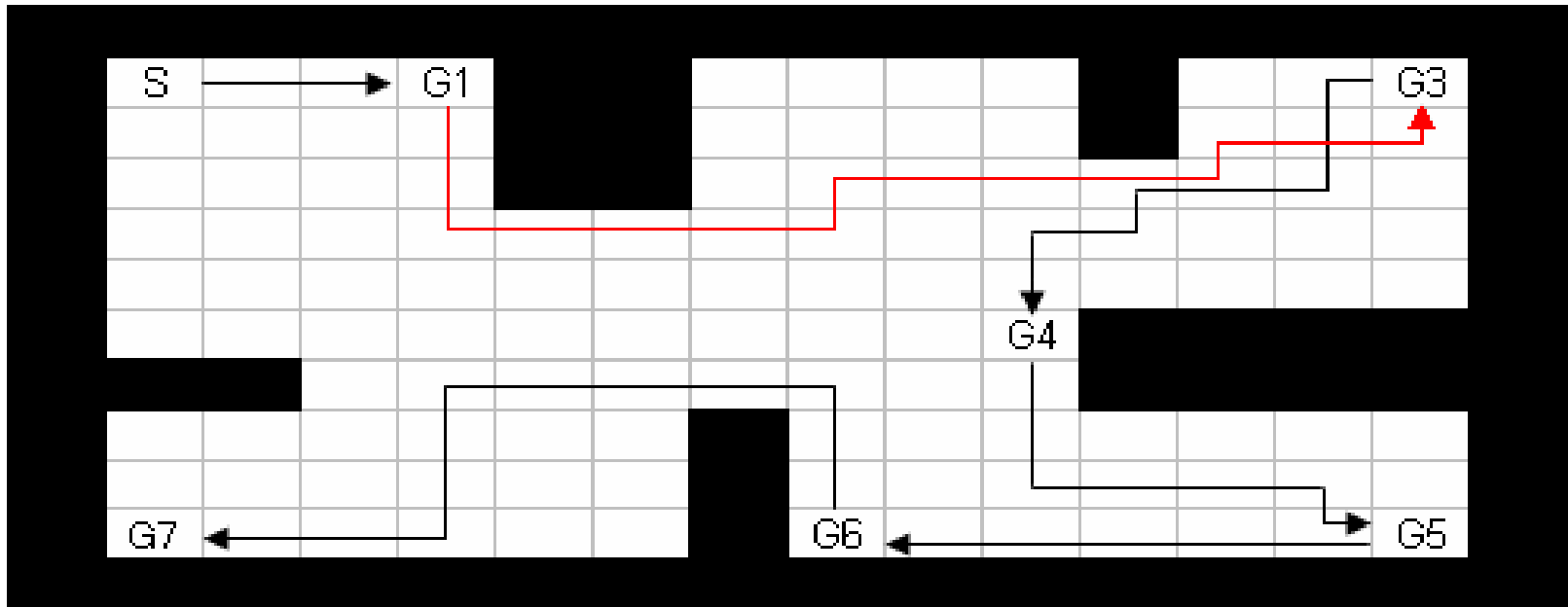
最短経路の行動回数  $n$

# 提案手法2の結果



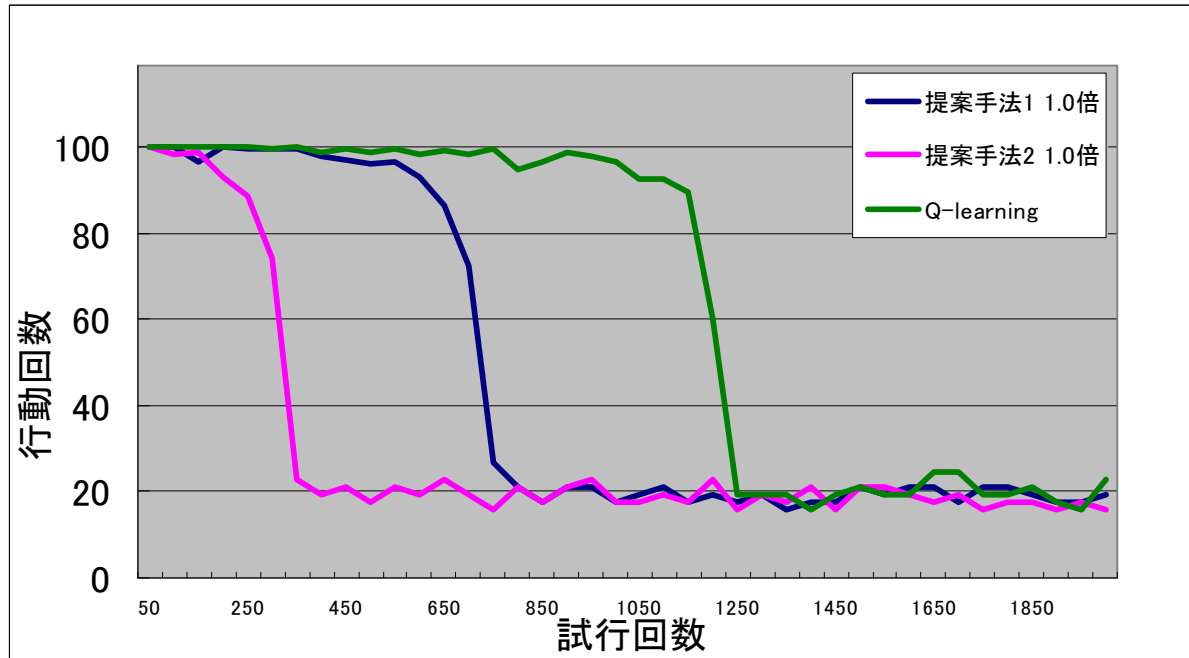
- Q-learningに比べ提案手法2は少ない試行回数で学習できた
- 提案手法の0.3～1.0倍は、ほぼ同じ結果となった

# ごみ削除実験結果



- 赤い矢印が学習した経路である
- G2を削除した場合でも最短経路の1つを学習できた

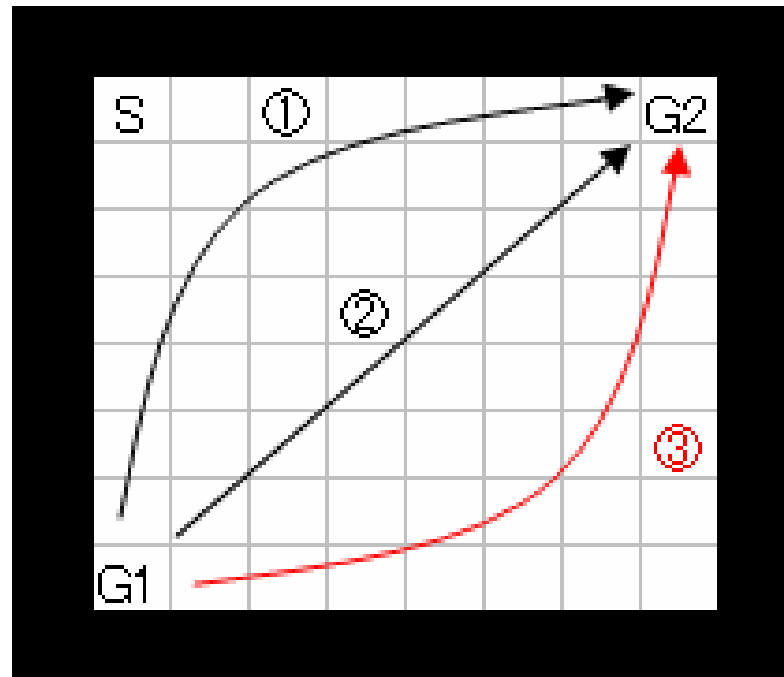
# 行動回数の推移



- グラフは各手法における行動回数の推移を比較したものである
- 提案手法を使うことで、少ない試行回数で学習ができていることが分かる

# 問題点

- 提案手法2では、図のG1を削除した時、経路①、②で学習していた場合、最短経路を学習できたが、経路③では最短経路を学習できなかった



# 今後の課題

- 提案手法1、提案手法2を組み合わせ問題点を解決できる条件付けを考える
- ごみの追加をする際にも学習したQ値を用いて学習効率を上げる方法を考える

# まとめ

- ごみを収集する最短経路を強化学習を用いて学習することができた
- 提案した手法を使えばカーナビ等の経路探索に用いられているルート案内に応用できるのではないかと考えられる