H29 [1]

(1) 定義に従って動的計画法により図3を埋めると以下の結果が得られる

−1	0	0
3	0	0
3	2	0

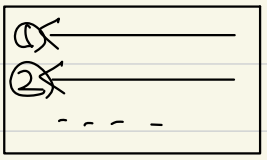
"

(2) (1)の動的計画法の結果を用いてテーブルを復元することで、最適な経路は以下のようになる.

"

(3) $V(S) = \max_{a \in (UP, DOWN)} \left\{ R_{ss'}^{a} + V(s') \right\}$ の定義に基いて

迷路の各マスの $V(S)$ の値を上の行の右のマスから動的計算により求めることを行えばよい.

(図示すると、更新の向きは右図の通り 

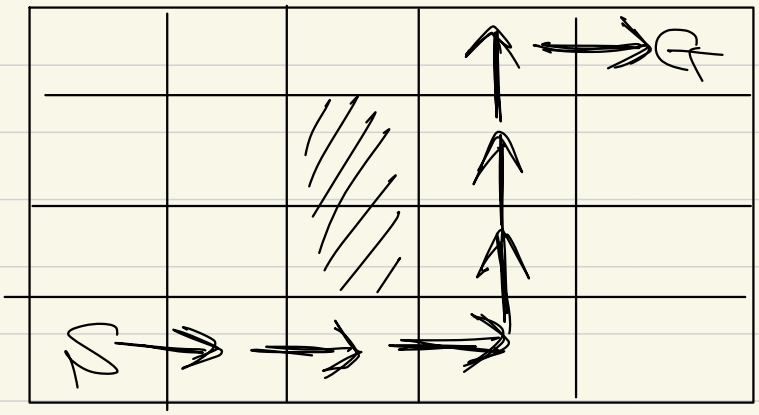
)

これにおて得られる $V(S)$ の値をテーブルにまとめると、次ページのように表現できる.

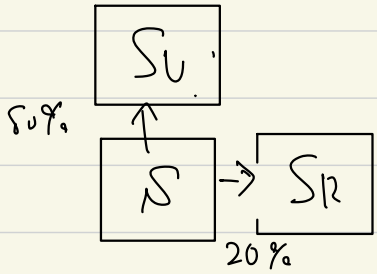
2	0	0	0	0
4	2	////	0	0
6	4	////	10	-5
10	10	10	10	-10

(通行不能なマスは //// で表記)

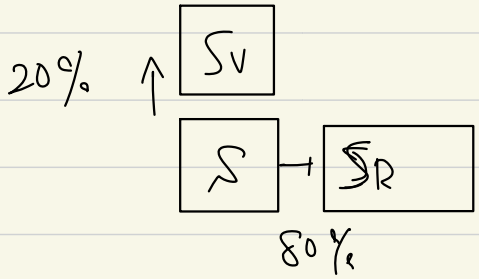
(4) (3)の動的計画法テーブルのデータから得られる最大利益値は10であり、それを与えるパスを復元するため始点Sから終点Gに向けてテーブルを戻ると、最適経路は以下のように復元できる。



以下、今いるマスを P とした時、その上のマスを P_U、その右のマスを P_R と定義する。また、便宜上障害物や迷路外にもアクセスできるものとし、そのような位置 x において、$V(x) = -\infty$ とする。



① 上を選択



② 右を選択

(但し、どちらかが不正なルートであった場合残りの方向へ確率1で移動する)

この時、$V(S)$ を計算するにあたって、進む方向として上右を選択するかに依存し①②のパターンがありうる

①を選択した場合

$$V(S) = \frac{4}{5}\left\{ R_{S.SU}^{a} + V(SU) \right\} + \frac{1}{5}\left\{ R_{S.SR}^{a} + V(SR) \right\}$$

②を選択した場合

$$V(S) = \frac{1}{5}\left\{ R_{S.SU}^{a} + V(SU) \right\} + \frac{4}{5}\left\{ R_{S.SR}^{a} + V(SR) \right\}$$

という関係式が成り立つ.

次にどちらにも行くことが可能である場合

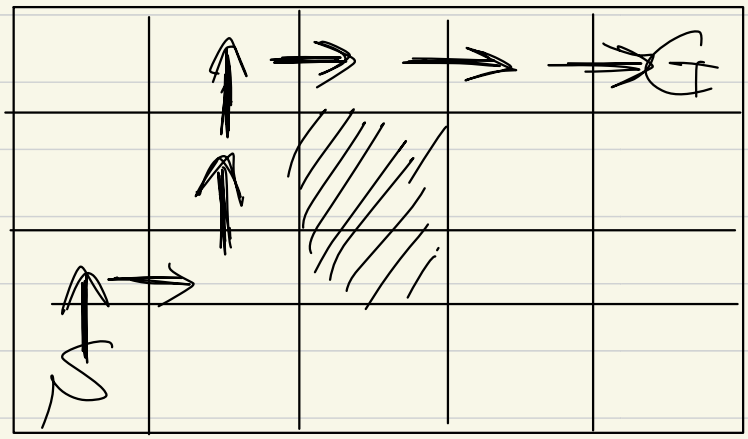
$$V(S) = \max\left\{ \frac{4}{5}\left(R_{S.SU}^{c} + V(SU) \right) + \frac{1}{5}\left(R_{S.SR}^{a} + V(SR) \right),\ \frac{1}{5}\left(R_{S.SU}^{a} + V(SU) \right) + \frac{4}{5}\left(R_{S.SR}^{c} + V(SR) \right) \right\}$$

//

(6) (5)で定めた仮定と、それにおいて得られる漸化式に基き（そうでなければ(1)～(4)と同じ）
(3)での更新順序と同じ順序で動的にテーブルを更新
すると、以下の結果を得る。（但し結果は小数第1位までとしている）

2	0	0	0	0
3.2	2	////	-1	0
5.8	4	////	5.2	-5
7.2	5.0	1.1	1.1	-10

(7) (4)と同様の方法で最適指示路を復元すると、//次ページのようになる.



(8) (4)で与えられる最適経路は、それが極めて限定された経路においてしか達成しえず、そこから少しでも外れてしまえば得られる報酬値は急速に悪化する。

一方、(7)で与えた最適経路では、複数の最大報酬を得ることのできる経路が存在するため、仮に経路が指示から外れてしまっても解の劣化が起こりにくい。

その差異が、得られる報酬の期待値として表れたことと考えられる。