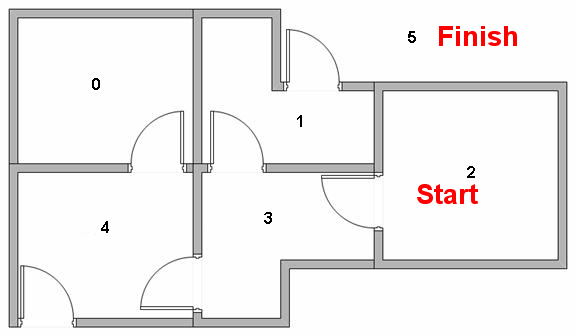
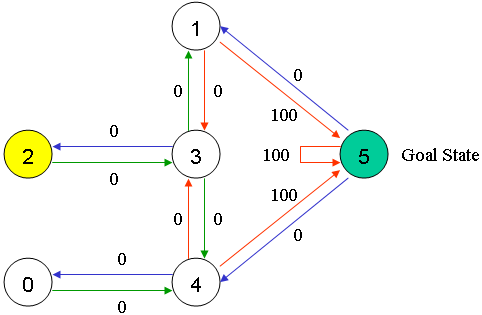
**Q-LEARNING**

**Bài toán:** Giả sử chúng ta có 5 phòng trong 1 căn nhà, các phòng được kết nối với nhau bằng các cửa được mô tả như hình dưới đây. Các căn phòng được đánh số từ 1-4, ngoài trời được đánh là số (5). Bạn đang đứng ở 1 căn phòng bất kì trong căn nhà đó, bạn phải di chuyển như thế nào để có thể ra ngoài nhanh nhất có thể. Sử dụng thuật toán Q-learning giải quyết vấn đề.



1. **Cơ sở lý thuyết**

Ta phác họa bài toán lại, với các hành động dẫn tới “Finish” sẽ được một “món quà” 100, các hành động còn lại xem như bằng 0.



Theo như hình trên, ta biết được các khả năng có thể thực hiện, khi ta đang đứng ở bất kì phòng nào (vì chúng được nối với nhau cố định từ trước). Từ đó, chúng ta quy định một ma trận trạng thái ban đầu, để đặc trưng cho khả năng thực hiện: ma trận R.



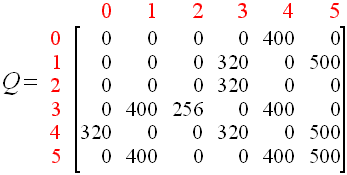
(với số -1 có nghĩa là không thể thực hiện khả năng đó)

Khởi tạo 1 ma trận Q, khi học xong thuật toán Q-learning, ma trận Q sẽ cho chúng ta biết đường đi tối ưu khi đang ở một căn phòng bất kì.

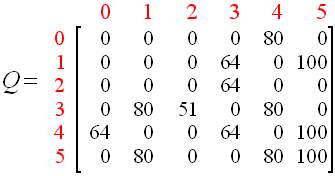
Thuật toán Q-learning được mô tả ngắn gọn như sau:

Q(state, action) 🡨 R(state, action) + γ\*Max[Q(next state, all actions)]

Sau một loạt lần lặp, ma trận Q được cập nhật tối ưu nhất

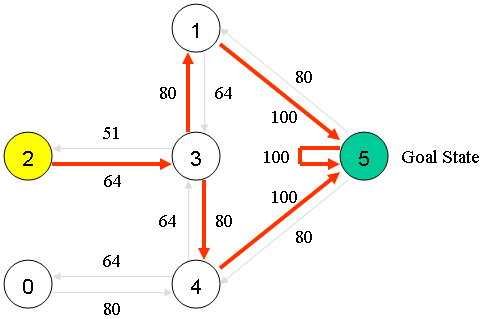


Chuẩn hóa lại ma trận Q:



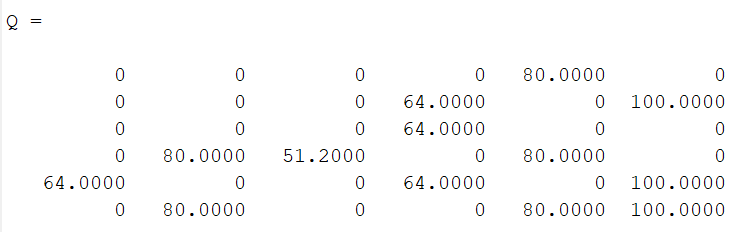
Theo như ma trận Q, ta có thể tìm được đường đi tối ưu khi đang ở một phòng bất kì. Ví dụ, ta đang ở phòng 3, nhìn vào ma trận, ta chọn đường đi nào có "phần thưởng" cao nhất --> chọn đi qua phòng 1 hoặc 4.

Từ đó, ta phác họa lại kết quả tối ưu cho bài toán ban đầu:



1. **Code Matlab** (trong file đính kèm)
2. **Kết quả Matlab**

Kết quả phù hợp với tính toán lý thuyết



Tuy nhiên, cần lặp với số vòng lặp lớn để tìm ra kết quả tối ưu. Nhóm em có làm với phép lặp nhỏ hơn, gần như không thể đạt kết quả ma trận Q mong muốn, nhưng kết quả có thể chấp nhận được.

