Practical Part 1. An Exercise in Dec-POMDP Problem Specification

1. States, Actions, and Observation

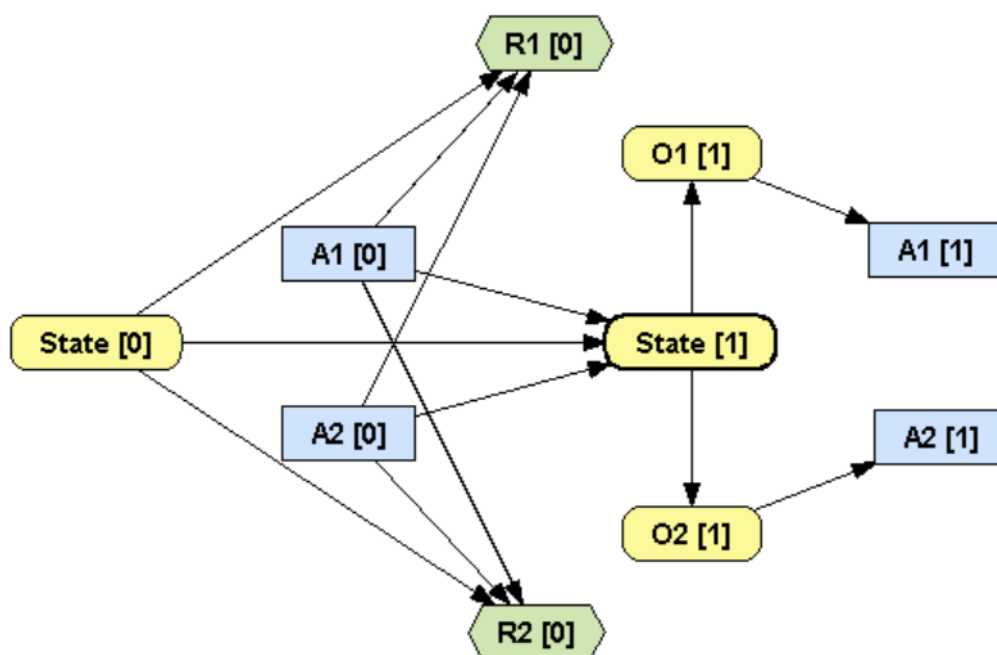a) $S = \{Tiger_{Left}, Tiger_{Right}\}$. No.

b)
$A = \{(Open_{Left}, Open_{Left}), (Open_{Left}, Open_{Right}), (Open_{Left}, Listen), (Open_{Right}, Open_{Left}), (Open_{Right}, Open_{Right}), (Open_{Right}, Listen), (Listen, Open_{Left}), (Listen, Open_{Right}), (Listen, Listen)\}$
Individual action space for agent i: $A_i = \{Open_{Left}, Open_{Right}, Listen\}$.

c)
$O = \{(Roar_{Left}, Roar_{Left}), (Roar_{Left}, Roar_{Right}), (Roar_{Right}, Roar_{Left}), (Roar_{Right}, Roar_{Right})\}$
. Individual observation space for agent i: $O_i = \{Roar_{Left}, Roar_{Right}\}$.

2. A Graphical Representation of the Dec-POMDP

State [1]的Probability，除了以下两种情况，其余都是uniform: 0.5, 0.5

State [0]	Tiger_Right	Tiger_Left
A1 [0]	Listen	Listen
A2 [0]	Listen	Listen
Tiger_Left	0	1
Tiger_Right	1	0

Observation:

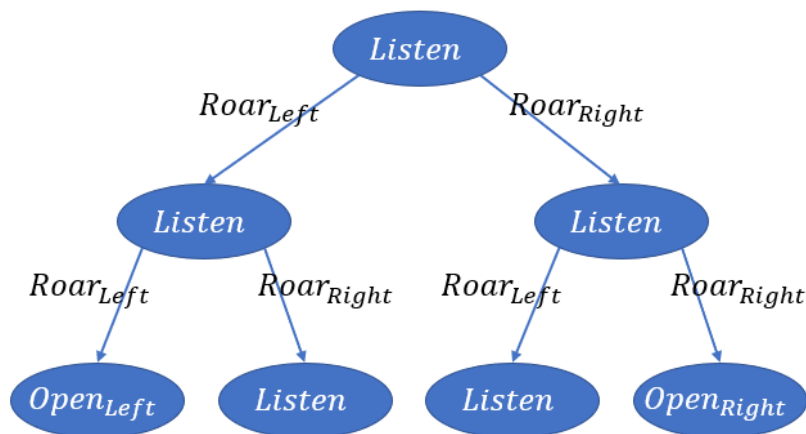| State [1]	Tiger_Right	Tiger_Left
Roar_Left	0.15	0.85
Roar_Right	0.85	0.15

Reward: （以Agent1为例）

State [0]	Tiger_Right	Tiger_Right	Tiger_Right	Tiger_Right	Tiger_Right	Tiger_Right	Tiger_Right	Tiger_Right	Tiger_Right
A1 [0]	Listen	Listen	Listen	Open_Right	Open_Right	Open_Right	Open_Left	Open_Left	Open_Left
A2 [0]	Listen	Open_Right	Open_Left	Listen	Open_Right	Open_Left	Listen	Open_Right	Open_Left
R1 [0]	−1	−1	−1	−100	−50	−100	10	0	20

Tiger_Left	Tiger_Left	Tiger_Left	Tiger_Left	Tiger_Left	Tiger_Left	Tiger_Left	Tiger_Left	Tiger_Left
Listen	Listen	Listen	Open_Right	Open_Right	Open_Right	Open_Left	Open_Left	Open_Left
Listen	Open_Right	Open_Left	Listen	Open_Right	Open_Left	Listen	Open_Right	Open_Left
−1	−1	−1	10	20	0	−100	−100	−50

见Dec-Tiger.pgmx

3. Policies

a) An arbitrary policy:

b) h=1: 两个Agent一直Listen

h=2: 两个Agent一直Listen

h=3: 两个Agent一直Listen（可能情况太多，不确定）

GMAA解得最优为：

```
Policy for agent 0 (index 55):
() --> listen
(hear-left) --> listen
(hear-right) --> listen
(hear-left,hear-left) --> open-right
(hear-left,hear-right) --> listen
(hear-right,hear-left) --> listen
(hear-right,hear-right) --> open-left
Policy for agent 1 (index 55):
() --> listen
(hear-left) --> listen
(hear-right) --> listen
(hear-left,hear-left) --> open-right
(hear-left,hear-right) --> listen
(hear-right,hear-left) --> listen
(hear-right,hear-right) --> open-left
OptimalValueDatabase: entry 'dectiger 1 3'
```

BFS解得最优为：

```
Policy for agent 0 (index 0):
Oempty,    --> listen
Oempty, hear-left,   --> listen
Oempty, hear-right,   --> listen
Oempty, hear-left, hear-left,   --> listen
Oempty, hear-left, hear-right,   --> listen
Oempty, hear-right, hear-left,   --> listen
Oempty, hear-right, hear-right,   --> listen
Policy for agent 1 (index 0):
Oempty,    --> listen
Oempty, hear-left,   --> listen
Oempty, hear-right,   --> listen
Oempty, hear-left, hear-left,   --> listen
Oempty, hear-left, hear-right,   --> listen
Oempty, hear-right, hear-left,   --> listen
Oempty, hear-right, hear-right,   --> listen
```

c) $|A_{max}|^{n \frac{|\Omega_{max}|^T - 1}{|\Omega_{max}| - 1}} = 3^{2 \times \frac{2^3 - 1}{2 - 1}} = 4,782,969$

4. Solution Methods.

a) h=1:

```
Sampled value = -2 (computed was -2)
====================== GMAA run 1/1 ended, Dec-POMDP value=-2

Summary of timing results:
ComputeQ: 0 s in 1 measurements, max 0, avg 0, min 0
Overall: 0.11 s in 1 measurements, max 0.11, avg 0.11, min 0.11
Plan: 0 s in 1 measurements, max 0, avg 0, min 0
PlanningUnit: 0 s in 1 measurements, max 0, avg 0, min 0
Simulation: 0.1 s in 1 measurements, max 0.1, avg 0.1, min 0.1
GMAA::Plan: 0 s in 1 measurements, max 0, avg 0, min 0
GMAA::Plan::iteration: 0 s in 1 measurements, max 0, avg 0, min 0
GMAA_MAAstar::NextExact_ts0: 0 s in 1 measurements, max 0, avg 0, min 0
evaluateRandomPolicy sampled value: -47.385
```

h=2:

```
Sampled value = -4 (computed was -4)
====================== GMAA run 1/1 ended, Dec-POMDP value=-4

Summary of timing results:
ComputeQ: 0 s in 1 measurements, max 0, avg 0, min 0
Overall: 0.11 s in 1 measurements, max 0.11, avg 0.11, min 0.11
Plan: 0 s in 1 measurements, max 0, avg 0, min 0
PlanningUnit: 0 s in 1 measurements, max 0, avg 0, min 0
Simulation: 0.1 s in 1 measurements, max 0.1, avg 0.1, min 0.1
GMAA::Plan: 0 s in 1 measurements, max 0, avg 0, min 0
GMAA::Plan::iteration: 0 s in 12 measurements, max 0, avg 0, min 0
GMAA_MAAstar::NextExact_ts0: 0 s in 9 measurements, max 0, avg 0, min 0
GMAA_MAAstar::NextExact_ts1: 0 s in 3 measurements, max 0, avg 0, min 0
evaluateRandomPolicy sampled value: -93.045
```

h=3:

```
Sampled value = 4.5254 (computed was 5.19081)
===================== GMAA run 1/1 ended, Dec-POMDP value=5.19081

Summary of timing results:
ComputeQ: 0 s in 1 measurements, max 0, avg 0, min 0
Overall: 0.22 s in 1 measurements, max 0.22, avg 0.22, min 0.22
Plan: 0.09 s in 1 measurements, max 0.09, avg 0.09, min 0.09
PlanningUnit: 0 s in 1 measurements, max 0, avg 0, min 0
Simulation: 0.12 s in 1 measurements, max 0.12, avg 0.12, min 0.12
GMAA::Plan: 0.09 s in 1 measurements, max 0.09, avg 0.09, min 0.09
GMAA::Plan::iteration: 0.09 s in 44 measurements, max 0.01, avg 0.00204545, min 0
GMAA_MAAstar::NextExact_ts0: 0 s in 9 measurements, max 0, avg 0, min 0
GMAA_MAAstar::NextExact_ts1: 0 s in 19 measurements, max 0, avg 0, min 0
GMAA_MAAstar::NextExact_ts2: 0.09 s in 16 measurements, max 0.01, avg 0.005625, min 0
evaluateRandomPolicy sampled value: -140.81
```

b) 如BFS：

h=1：

```
real    0m0.032s
user    0m0.014s
sys     0m0.021s
```

```
value=-2
JointPolicyPureVector:
JPolComponent_VectorImplementation index 0
Policy for agent 0 (index 0):
Oempty,   --> listen
Policy for agent 1 (index 0):
Oempty,   --> listen
```

h=2：

```
real    0m0.047s
user    0m0.030s
sys     0m0.005s
```

```
value=8
JointPolicyPureVector:
JPolComponent_VectorImplementation index 0
Policy for agent 0 (index 0):
Oempty,   --> listen
Oempty, hear-left,   --> listen
Oempty, hear-right,   --> listen
Policy for agent 1 (index 0):
Oempty,   --> listen
Oempty, hear-left,   --> listen
Oempty, hear-right,   --> listen
```

h=3：

```
real    0m42.689s
user    0m42.229s
sys     0m0.049s
```

```
value=6
JointPolicyPureVector:
JPolComponent_VectorImplementation index 0
Policy for agent 0 (index 0):
Oempty,  --> listen
Oempty, hear-left,  --> listen
Oempty, hear-right,  --> listen
Oempty, hear-left, hear-left,  --> listen
Oempty, hear-left, hear-right,  --> listen
Oempty, hear-right, hear-left,  --> listen
Oempty, hear-right, hear-right,  --> listen
Policy for agent 1 (index 0):
Oempty,  --> listen
Oempty, hear-left,  --> listen
Oempty, hear-right,  --> listen
Oempty, hear-left, hear-left,  --> listen
Oempty, hear-left, hear-right,  --> listen
Oempty, hear-right, hear-left,  --> listen
Oempty, hear-right, hear-right,  --> listen
```

(内容合理即可)