Báo cáo của các bạn nên được trình bày với các nội dung như sau:

# Thông tin nhóm: MSSV, họ và tên các thành viên trong nhóm (theo thứ tự MSSV).

19120685 – Võ Ngọc Tín

# I - Phát biểu bài toán: mô tả các bài toán cần giải quyết.

## Sơ lược về bài toán crying baby

A partially observable Markov decision process (POMDP) – quy trình quyết định Markov có thể quan sát được một phần

Bài toán “The crying baby” là một dạng POMDP đơn giản với 2 trạng thái (states), 3 hành động (actions) và 2 trạng thái quan sát (observations).

Mục đích của bài toán là làm thế nào để chăm sóc một em bé tốt, bằng cách là lựa chọn thời gian nào thích hợp để cho ăn (feed), hát (sing) hay có thể ngó lơ (ignore) em bé.

2 trạng thái (states) gồm: đói (hungry) và no (stated)

3 hành động (actions) gồm: cho ăn (feed), hát (sing) và ngó lơ (ignore)

2 trạng thái quan sát (observations) gồm: khóc (crying) và im lặng (quiet)

States = {hungry, stated}

Actions = {feed, sing, ignore}

Observations = {crying, quiet}

**TRANSITIONS**

Mối liên hệ giữa các hành động và trạng thái:

Hungry + feed => stated 100%

Hungry + sing => hungry 100%

Hungry + ignore => hungry 100%

Stated + feed => stated 100%

Stated + sing => hungry 10%

Stated + ignore => hungry 10%

**OBSERVATIONS**

Mối liên hệ giữa trạng thái, hành động và trạng thái quan sát được

Hungry + feed => cry 80%

Hungry + sing => cry 90%

Hungry + ignore => cry 90%

Stated + feed => cry 10%

Stated + sing => cry 0%

Stated + ignore => cry 10%

Lưu ý: cry + quiet = 100% =>

Cry = 10% ⬄ quiet = 90%

**REWARD FUNCTION**

Hungry độc lập với các actions khác => -10

Feed => -5

Sing => -0.5

## Bài toán multicare giver crying baby

Consider a multiagent POMG generalization of the crying baby problem.

We have two caregivers taking care of a baby. As in the POMDP version, the states are the baby being hungry or sated. Each caregiver’s actions are to feed, sing, or ignore the baby. If both caregivers choose to perform the same action, the cost is halved. For example, if both caregivers feed the baby, then the reward is only −2.5 instead of −5. However, the caregivers do not perfectly

observe the state of the baby. Instead, they rely on the noisy observations of the baby crying, both with the same observation. As a consequence of

the reward structure, there is a trade-off between helping each other and greedily choosing a less costly action.

Xem xét một POMG tổng quát đa phương pháp về vấn đề trẻ khóc.

Chúng tôi có hai người chăm sóc chăm sóc một em bé. Như trong phiên bản POMDP, các trạng thái là em bé đang đói hoặc ăn. Mỗi hành động của người chăm sóc là cho trẻ ăn, hát hoặc phớt lờ em bé. Nếu cả hai người chăm sóc chọn thực hiện cùng một hành động, chi phí sẽ giảm một nửa. Ví dụ, nếu cả hai người chăm sóc đều cho em bé ăn, thì phần thưởng chỉ là −2,5 thay vì −5. Tuy nhiên, những người chăm sóc không hoàn toàn

quan sát tình trạng của em bé. Thay vào đó, họ dựa vào những quan sát ồn ào của đứa trẻ khóc, cả hai đều có cùng quan sát. Như một hệ quả của

cơ cấu phần thưởng, có sự đánh đổi giữa việc giúp đỡ lẫn nhau và việc tham lam lựa chọn một hành động ít tốn kém hơn.

Bài toán multi care giver crying baby là một dạng mở rộng của bài toán crying baby phía trên.

Bài toán này thuộc POMG - Partially Observable Markov Games – có thể được xem là phần mở rộng của POMDP với nhiều agent (ở bài toán này là Caregiver)

Cụ thể, mỗi caregiver I trong tập I = {1, 2} có states, actions và observations tương tự phía trên.

States = {hungry, stated}

Actions = {feed, sing, ignore}

Observations = {crying, quiet}

**TRANSITION, OBSERVATION DYNAMICS**

Tương tự như bài toán crying baby, ngoại trừ mỗi người caregiver

Hungry + (feed, \*) => stated 100%

Hungry + (\*, feed) => stated 100%

\*: tất cả các phép gán khác có thể có, trong dấu () là hành động của 2 caregiver

Nếu action không phải feed

Hungry + (\*, \*) => hungry 100%

Stated + (\*, \*) => stated 50%

Các trạng thể động có thể quan sát tương tự như bài toán crying baby, tuy nhiên cả 2 caregiver sẽ tạo ra các trạng t thái quan sát giống nhau của baby nhưng không nhất thiết phải lựa chọn action của người còn lại

Hungry + (sing, \*) = hungry + (\*, sing) = (cry, cry) = 90%

Hungry + (sing, \*) = hungry + (\*, sing) = (quiet, quiet) = 10%

stated + (sing, \*) = stated + (\*, sing) = (cry, cry) = 0%

Nếu action không phải sing, thì observations như sau:

Hungry + (\*, \*) = hungry + (\*, \*) = (cry, cry) = 90%

Hungry + (\*, \*) = hungry + (\*, \*) = (quiet, quiet) = 10%

stated + (sing, \*) = stated + (\*, sing) = (cry, cry) = 0%

stated + (\*, \*) = stated + (\*, \*) = (quiet, quiet) = 100%

Cả 2 caregiver đều muốn giúp khi baby khóc: -10

Người thứ nhất sở trường là feed:

Người thứ hai sở trường là sing:

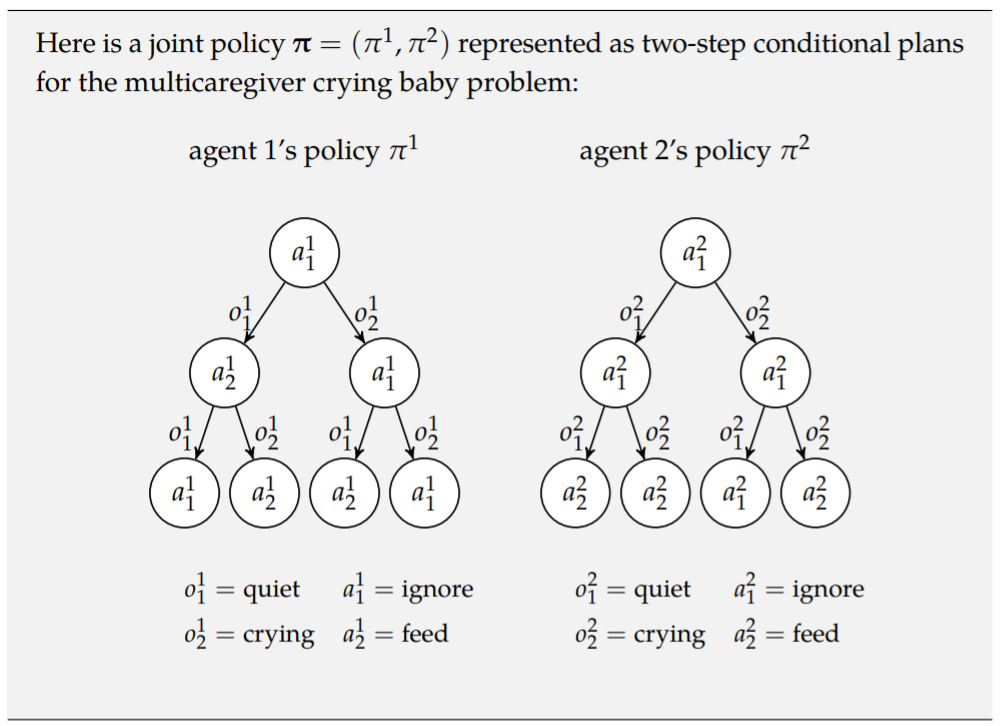
Khi feed, người thứ nhất -2.5, người thứ hai 5.0

Khi sing, người thứ nhất -0.5, người thứ hai -0.25

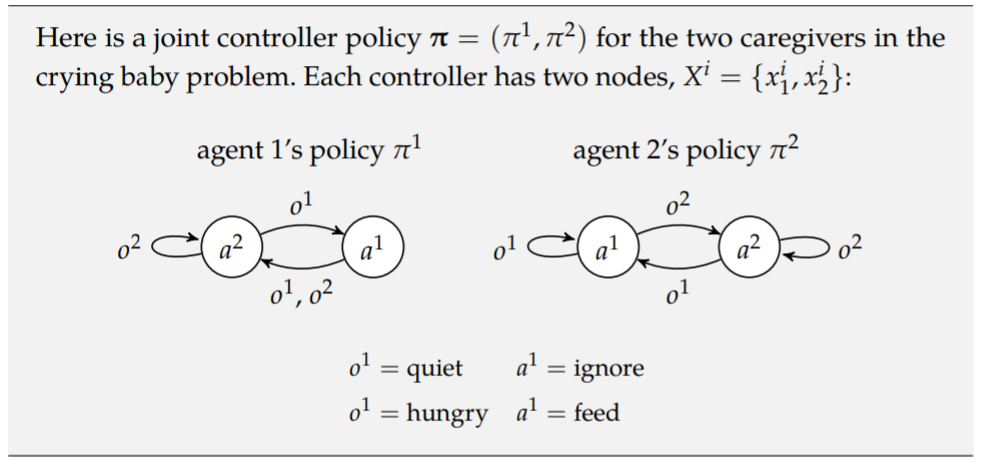
Text

Description automatically generated

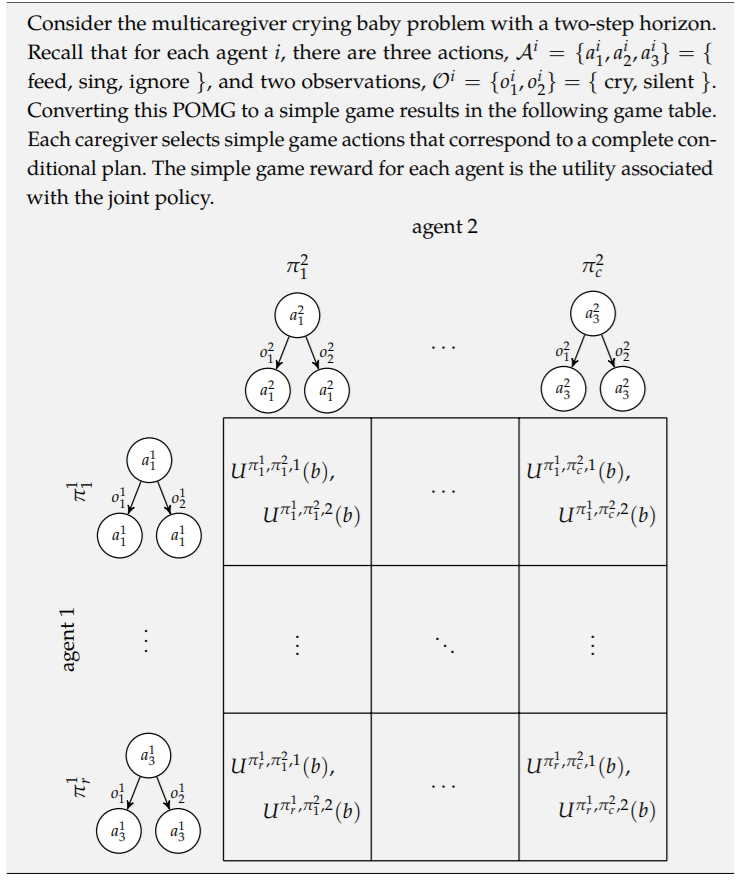
Example 26.2. A two-agent, twostep joint policy using conditional plans for the multicaregiver crying baby problem.



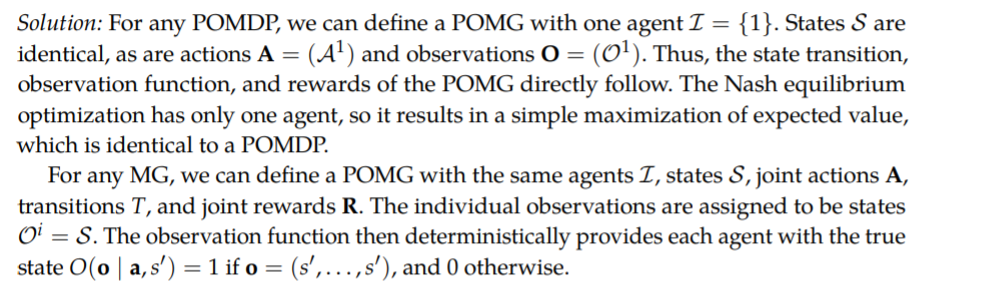
Example 26.3. A two-agent joint policy using controllers for the multicaregiver crying baby problem.



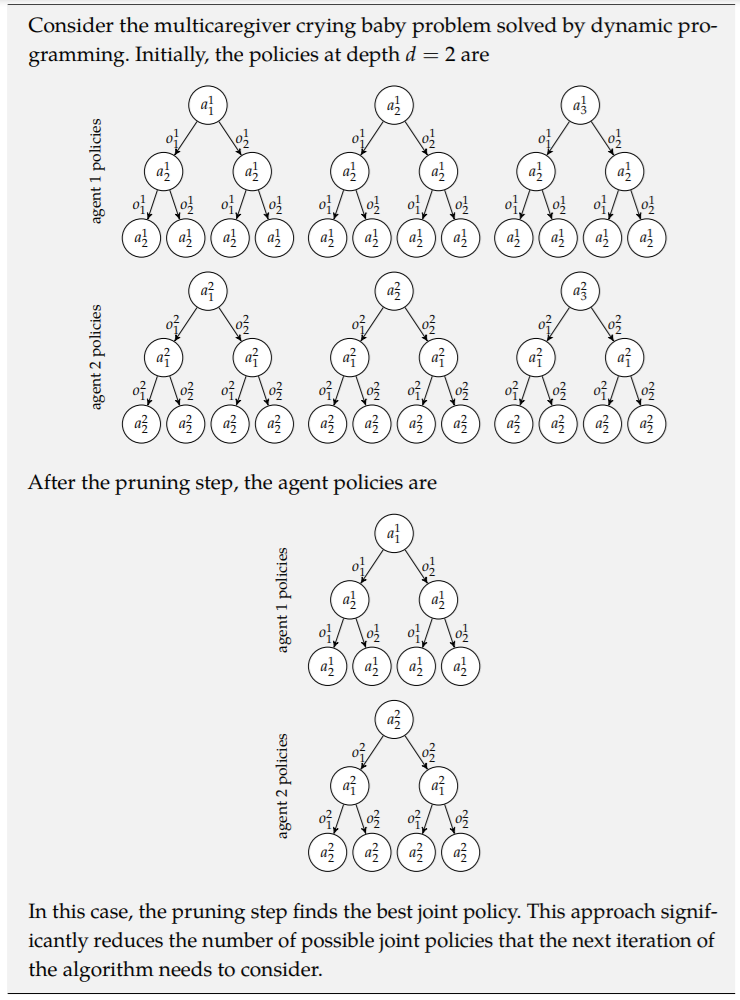
Example 26.4. Computing a Nash equilibrium for the multicaregiver crying baby problem by converting it into a simple game where the actions correspond to conditional plans.



Example 26.5. Dynamic programming and a single pruning step for the multicaregiver crying baby problem.



Solution: For any POMDP, we can define a POMG with one agent I = {1}. States S are identical, as are actions A = (A1 ) and observations O = (O1 ). Thus, the state transition, observation function, and rewards of the POMG directly follow. The Nash equilibrium optimization has only one agent, so it results in a simple maximization of expected value, which is identical to a POMDP. For any MG, we can define a POMG with the same agents I, states S, joint actions A, transitions T, and joint rewards R. The individual observations are assigned to be states Oi = S. The observation function then deterministically provides each agent with the true state O(o | a,s ′ ) = 1 if o = (s ′ , . . . ,s ′ ), and 0 otherwise.



# II - Thách thức: khó khăn, thách thức cụ thể đối với từng bài toán.

• POMGs generalize POMDPs to multiple agents and MGs to partial observability. • Because agents generally cannot maintain beliefs in POMGs, policies typically take the form of conditional plans or finite state controllers. • Nash equilibria, in the form of d-step conditional plans for POMGs, can be obtained by finding Nash equilibria for simple games whose joint actions consist of all possible POMG joint policies. • Dynamic programming approaches can be used to compute Nash equilibria more efficiently by iteratively constructing sets of deeper conditional plans while pruning dominated plans to restrict the search space.

# III - Thực nghiệm:

o Nêu rõ cấu hình các bài toán thực nghiệm.

o Với mỗi thử nghiệm:

 Mô hình hóa tính toán: cấu trúc dữ liệu hoặc phương pháp thể hiện các bài

toán.

 Phương pháp giải quyết: phương pháp, thuật toán được sử dụng cho việc giải

quyết bài toán, lý do lựa chọn phương pháp.

 Code: hướng dẫn và mô tả về đoạn code mà bạn đã viết (hay sử dụng). Lưu ý:

cần ghi rõ nguồn (nếu bạn sao chép đoạn code ở đâu đó).

 Phân tích: kết quả thử nghiệm và ý nghĩa.

# IV - Tóm tắt kết quả:

o Dựa trên các tiêu chí đánh giá kết quả của bạn.

o Điểm mạnh và điểm yếu trong đồ án này của bạn.