thymio & ai

ACTIVITY 2

The Arena

Reinforcement learning

Author
Joël RIVET

Graphic Designer
David COHEN

MOBSYA

ACTIVITY 2

The Arena

Reinforcement learning

Difficulty **Medium**

Estimated Time **60 min**

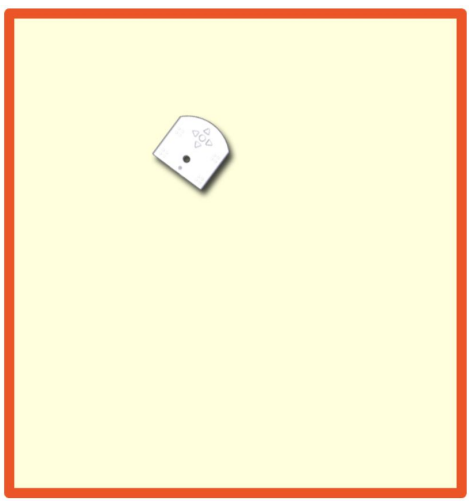
Prerequisites

Get to Know Thymio.

You have already **completed** the Thymio AI **First Learning activity**.

Material preparation

- Have a rectangular enclosure made of solid walls or a space delimited by objects heavy enough for Thymio not to move them.
- Minimum dimensions: 50cm x 50 cm.
- Arrange the Thymio robot inside the arena.



Teacher's manual

ACTIVITY 2

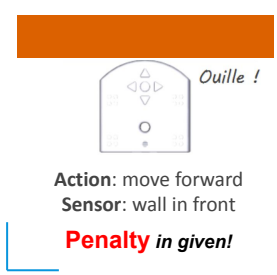
The Arena

Thymio's mission and its learning principle

- Thymio is inside an arena. Its goal is to explore this arena without touching the walls. As usual, at the beginning, Thymio does not know what or how to do.
- We decide to teach him to do his mission with a new method called **reinforcement learning.** This method is also part of the field of artificial intelligence.
- We will use again a neural network.
- In this method, we don't tell Thymio what action to do, among the various possible actions. It is up to the robot to choose one. But how can it know which action to perform? Well … through a rewards scoring it is given, or not[1].

Reinforcement learning

- Our new way of training Thymio will consist of telling it whether the action it has chosen, for a given sensor situation, is a good or bad choice by giving it a reward or a penalty for the action it is performing.
- The neural network is thus aware of these rewards for each action performed. How will the network react? Like a human being or an animal, it will try to obtain as many rewards as possible and to avoid penalties as much as possible.
- The network uses mathematics to progress. If one doubted the usefulness of mathematics, here is a good example of its application.
- Step by step, Thymio will perform more and more the **"good "** actions (those that bring rewards), and less and less the **"bad "** ones.
- To discover the good or bad actions, the robot must explore all of them.
- So, from time to time, the robot will try new actions, regardless of the rewards.



Teacher's manual

[1] This remind us the guessing game, You're cold, you're hot, you're burning!

Alphai Settings

Reminder on how to connect Thymio

1. **Connect Thymio** with one of the following 2 methods:
   - either connect the robot to the computer using the USB cable
   - or connect the USB key and turn on the robot
2. **Launch** the **Thymio Suite software** and wait a few moments.
3. **Launch** the **Alphai software**.
4. In **Alphai**, select your **robot** that appears on the blue screen to **establish the connection**.

Settings

- The settings are located in a file we will need to load.
- In the **settings** menu, select the option **load sample settings**...
- In the box that appears, double-click on the option **obstacle avoidance** (reinforcement learning).
- Under the network you will notice two black progress bars.
  - The left bar displays the rewards or penalties as a number.
    A reward is represented by a positive number, a negative value corresponds to a penalty. The possible values for the rewards are already set by Alphai.
  - The right bar indicates the level, i.e. the average of the previous rewards. It corresponds to the state of learning of Thymio.
- Besides, we will leave the **learning** and **exploring** buttons activated.

ACTIVITY 2

A first training experience

- **Observe Thymio in the arena**
- Launch the **self-drive** mode. Thymio starts to move. Its first move is chosen randomly. Observe its movements and color changes. Click the **self-drive** button again after about 20 seconds to stop the learning process.

Question 1: Draw associations between what the robot is doing and the colors it is using. Write down the answer:

When Thymio moves forward in the arena, it is green. When it gets too close or touches a wall, it turns red.

Green means: Thymio receives a reward, its action is in line with its mission.

Red means: Thymio receives a penalty, its action is contrary to its mission.

Continue the training for 5-10 minutes.

Question 2: Do you notice a change in Thymio's behavior? If so, describe it. Write down the answer:

We observe that at the beginning, Thymio hits the walls quite often and struggles to get away from them.

Then it manages to avoid them more and more often. We can even notice several stages in the learning process:

- At the beginning Thymio discovered very quickly that it was not necessary to go backwards.
- Then it quickly adopted one of two behaviors (students will observe one or the other with their Thymio): either turning in circles (this is the most common), or making straight lines and turning around when it hits a wall.
- Then it gradually learns to alternate between going straight and turning, choosing straight more and more frequently when there is no obstacle in front, turning left if there is a wall on the right, and turning right if there is a wall on the left.

The role of exploration

- Click on the **exploration** button to deactivate it.

Question 3: Do you notice an evolution in Thymio's behavior?

If so, describe it. Write down your answer:

Thymio makes fewer mistakes, it no longer interrupts its straight lines with unexpected movements.

Reset the AI with the **reset AI (reset learning) button**. This makes Thymio forget everything it previously learnt. It starts learning again from scratch (keep **self-drive** enabled, but **exploration** disabled).

Question 4: Do you notice any differences between this new learning and the previous one? If so, describe them. If you do not notice any difference, start a new learning process by pressing **reset AI (reset learning)** .

Write down your answer:

Thymio gets stuck in the “going in circles" behavior without discovering the straight line (This is not systematic and it sometimes learns correctly even when exploration is disabled).

Conclusion

Exploration is essential for learning.

- The AI occasionally tries actions other than the one it “thinks" is the best (when this happens the action icon on the right of the screen lights up in blue instead of black). This avoids getting stuck in a poor behavior.
- On the other hand, once learning is complete, exploration is no longer useful, so it's worth turning it off to get the most perfect behavior possible."

ACTIVITY 2

The behavior of the neural network

Observe the behavior of the neural network

We will observe and note the behavior in detail over a few steps, at the beginning of the learning process.

1. Reset the AI with the **reset AI (reset learning)** button. Place Thymio in the middle of the arena. Remember that its first move is chosen randomly. To be sure, you can click several times in a row on **reset AI (reset learning)** and **self-drive**.
2. Complete the first line of the table. The small dash means that the front sensors of Thymio do not detect anything because there is nothing.
3. Watch Thymio carefully and click on the **step by step** button.

Step by Step phase	Front sensors	Move backward to the left	Move forward to the left	Move forward	Move backward to the right	Move forward to the right	Reward value	Level value
								

ACTIVITY 2

The Arena

Question 5: What movement did Thymio make? In the table-row you just filled in, write down the value corresponding to this movement. Compare this number to the values of the other actions. What do you notice?

Write down the answer:

This value is the highest.

It means, Thymio performs the movement corresponding to the highest value.

Question 6: We also observe that the robot has received a reward. Does this reward seem to be consistent with the purpose of the mission?

Explain:

Possible answer: Thymio turned left, an it received a reward of +55. This is normal, because there is nothing in front of him, so it can make the turn.

Once the first reward is given, the output values are recalculated by the neural network. Complete the 2nd line of the table and guess what will be the next movement of Thymio.

Make a few more clicks on the **step by step** button while observing the evolution of the rewards and the level.

ACTIVITY 2

The role of rewards

Observe the behavior of the neural network

We will observe and note the behavior in detail over a few steps, at the beginning of the learning process.

1. Reset the AI with the **reset AI (reset learning)** button. Place Thymio in the middle of the arena. Remember that its first move is chosen randomly. To be sure, you can click several times in a row on **reset AI (reset learning)** and **self-drive**.
2. Complete the first line of the table. The small dash means that the front sensors of Thymio do not detect anything because there is nothing.
3. Watch Thymio carefully and click on the **step by step** button.

Step by Step phase	Front sensors	Move backward to the left 	Move forward to the left 	Move forward 	Move backward to the right 	Move forward to the right 	Reward value	Level value

ACTIVITY 2

Question 7: How does the level change when the robot receives a reward or on the contrary a penalty? What does the level represent?

Write down the answer:

- If the reward is positive, the level increases.
- If the reward is negative, the level decreases.

The level represents Thymio's ability to obtain positive rewards. Specifically, it is calculated as the average of the rewards received during the past minute.

Now press **self-drive** to let Thymio continue learning.

Question 8: How does the level change during the learning process? Why?

Explain:

The level increases during the learning process. Indeed Thymio receives more and more high rewards (especially when it goes straight) and less and less punishments (since it hits itself less and less). In fact, the goal of learning is precisely to make Thymio's level increase.

ACTIVITY 2

The Arena

The influence of the penalty on taught behavior

If we summarize the different values that appeared in the progress bar, we noticed:

- **100:** When Thymio goes straight ahead without any obstacle in front, this is the highest value.
- **55:** When Thymio turns without obstacle in front of it.
- **-50:** Thymio performs one of the multiple “bad" actions like moving against a wall or moving backwards when there is nothing in front, etc.

We can change the value of the penalty.

- Open the reward tab and set the penalty to a small value, for example 0 : reset the AI and restart the learning for a few minutes.
  Observe Thymio's behavior, is it more daring or more cautious? Does it hit the walls more or less often?
- Set a bigger penalty, 1.5 for example. Reset the AI again and restart the learning process for a few minutes. Same question as before.

Question 9: Summarize how Thymio's behavior changes when you change the penalty value.

Write down the answer:

- **If the penalty is low,** Thymio often hits the walls but becomes bolder and explores the area fully.
- **If the penalty is high**, Thymio hits the walls less often but becomes more cautious and stays in a restricted zone.

ACTIVITY 2

The Arena

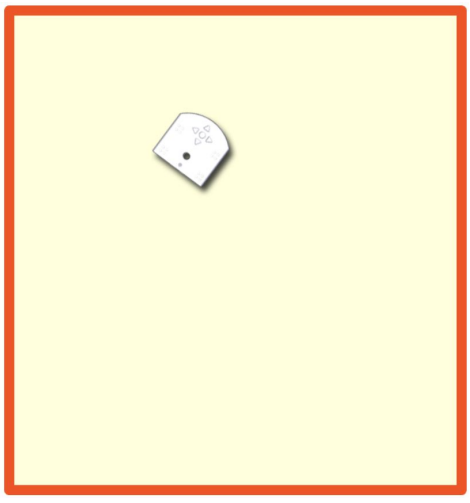
Reinforcement learning

Difficulty **Medium**

Estimated Time **60 min**

Thymio's mission and its learning principle

- Thymio is inside an arena. Its goal is to explore this arena without touching the walls. As usual, at the beginning, Thymio does not know what or how to do.
- We decide to teach him to do his mission with a new method called **reinforcement learning.** This method is also part of the field of artificial intelligence.
- We will use again a neural network.
- In this method, we don't tell Thymio what action to do, among the various possible actions. It is up to the robot to choose one. But how can it know which action to perform? Well … through a rewards scoring it is given, or not[1].

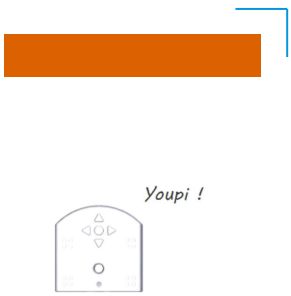


[1] This remind us the guessing game, You're cold, you're hot, you're burning!

ACTIVITY 2

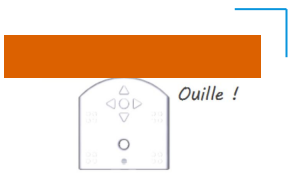
Reinforcement learning

- Our new way of training Thymio will consist of telling it whether the action it has chosen, for a given sensor situation, is a good or bad choice by giving it a reward or a penalty for the action it is performing.
- The neural network is thus aware of these rewards for each action performed. How will the network react? Like a human being or an animal, it will try to obtain as many rewards as possible and to avoid penalties as much as possible.
- The network uses mathematics to progress. If one doubted the usefulness of mathematics, here is a good example of its application.
- Step by step, Thymio will perform more and more the **“good "** actions (those that bring rewards), and less and less the **"bad "** ones.
- To discover the good or bad actions, the robot must explore all of them.
- So, from time to time, the robot will try new actions, regardless of the rewards.



Action: move forward
Sensor: nothing in front

Reward ***is granted!***



Action: move forward
Sensor: wall in front

Penalty ***in given!***

Alphai Settings

Reminder on how to connect Thymio

1. **Connect Thymio** with one of the following 2 methods:
   - either connect the robot to the computer using the USB cable
   - or connect the USB key and turn on the robot
2. **Launch** the **Thymio Suite software** and wait a few moments.
3. **Launch** the **Alphai software**.
4. In **Alphai**, select your **robot** that appears on the blue screen to **establish the connection**.

Settings

- The settings are located in a file we will need to load.
- In the **settings** menu, select the option **load sample settings**...
- In the box that appears, double-click on the option **obstacle avoidance** (reinforcement learning).
- Under the network you will notice two black progress bars.
  - The left bar displays the rewards or penalties as a number.
    A reward is represented by a positive number, a negative value corresponds to a penalty. The possible values for the rewards are already set by Alphai.
  - The right bar indicates the level, i.e. the average of the previous rewards. It corresponds to the state of learning of Thymio.
- Besides, we will leave the **learning** and **exploring** buttons activated.

ACTIVITY 2

A first training experience

- **Observe Thymio in the arena**
- Launch the **self-drive** mode. Thymio starts to move. Its first move is chosen randomly. Observe its movements and color changes. Click the **self-drive** button again after about 20 seconds to stop the learning process.

Question 1: Draw associations between what the robot is doing and the colors it is using. Write down your answer:

Continue the training for 5-10 minutes.

Question 2: Do you notice a change in Thymio's behavior? If so, describe it. write down the answer:

The role of exploration

- Click on the **exploration** button to deactivate it.

Question 3: Do you notice an evolution in Thymio's behavior?

If so, describe it. Write down your answer:

Reset the AI with the **reset AI (reset learning) button**. This makes Thymio forget everything it previously learnt. It starts learning again from scratch (keep **self-drive** enabled, but **exploration** disabled)

Question 4: Do you notice any differences between this new learning and the previous one? If so, describe them. If you do not notice any difference, start a new learning process by pressing **reset AI (reset learning)**.

Write down your answer:

ACTIVITY 2

The behavior of the neural network

Observe the behavior of the neural network

We will observe and note the behavior in detail over a few steps, at the beginning of the learning process.

1. Reset the AI with the **reset AI (reset learning)** button. Place Thymio in the middle of the arena. Remember that its first move is chosen randomly. To be sure, you can click several times in a row on **reset AI (reset learning)** and **self-drive**.
2. Complete the first line of the table. The small dash means that the front sensors of Thymio do not detect anything because there is nothing.
3. Watch Thymio carefully and click on the **step by step** button.

Step by Step phase	Front sensors	Move backward to the left	Move forward to the left	Move forward	Move backward to the right	Move forward to the right	Reward value	Level value
								

ACTIVITY 2

Question 5: What movement did Thymio make? In the table row you just filled in, write down the value corresponding to this movement. Compare this number to the values of the other actions. What do you notice?

Write down the answer:

Question 6: We also observe that the robot has received a reward. Does this reward seem to be consistent with the purpose of the mission?

Explain:

Once the first reward is given, the output values are recalculated by the neural network. Complete the 2nd line of the table and guess what will be the next movement of Thymio.

Make a few more clicks on the **step by step** button while observing the evolution of the rewards and the level."

Question 7: How does the level change when the robot receives a reward or on the contrary a penalty? What does the level represent?

Write down the answer:

Now press **self-drive** to let Thymio continue learning.

Question 8: How does the level change during the learning process? Why?

Explain:

ACTIVITY 2

The influence of the penalty on taught behavior

If we summarize the different values that appeared in the progress bar, we noticed:

- **100:** When Thymio goes straight ahead without any obstacle in front, this is the highest value.
- **55:** When Thymio turns without obstacle in front of it.
- **-50:** Thymio performs one of the multiple “bad" actions like moving against a wall or moving backwards when there is nothing in front, etc.

We can change the value of the penalty.

- Open the reward tab and set the penalty to a small value, for example 0 : reset the AI and restart the learning for a few minutes.
  Observe Thymio's behavior, is it more daring or more cautious? Does it hit the walls more or less often?
- Set a bigger penalty, 1.5 for example. Reset the AI again and restart the learning process for a few minutes. Same question as before.

Question 9: Summarize how Thymio's behavior changes when you change the penalty value.

Write down the answer:

ACTIVITY 2

The Arena

Answers

Name: ..... Class: .....

Question 1: Draw associations between what the robot is doing and the colors it is using.

Question 2: Do you notice a change in Thymio's behavior? If so, describe it

Question 3: Do you notice an evolution in Thymio's behavior? If so, describe it.

ACTIVITY 2

The Arena

Answers

Name: ..... Class: .....

Question 4: Do you notice any differences between this new learning and the previous one?

Question 5: What movement did Thymio make? In the table row you just filled in, write down the value corresponding to this movement. Compare this number to the values of the other actions. What do you notice?

Question 6: We also observe that the robot has received a reward. Does this reward seem to be consistent with the purpose of the mission? Complete the 2nd line of the table and guess what will be the next movement of Thymio.

ACTIVITY 2

The Arena

Answers

Name: ..... Class: .....

Question 7: How does the level change when the robot receives a reward or on the contrary a penalty? What does the level represent?

Question 8: How does the level change during the learning process? Why?

Question 9: Summarize how Thymio's behavior changes when you change the penalty value.

The Arena

The Mission

- **Thymio** is inside a **closed arena**.

- Its **goal** is to **explore** this **arena without touching the walls**.

Reinforcement Learning

- Each time **Thymio performs** an **action**, it is **told** whether the action it did, for a **given sensor situation**, is a **good** or **bad choice** through **rewards** or **penalties** scoring.

- **Thymio** will choose its **actions** in order to **get the maximum rewards**.

Action: move forward
Sensor: nothing in front
Reward ***is granted!***

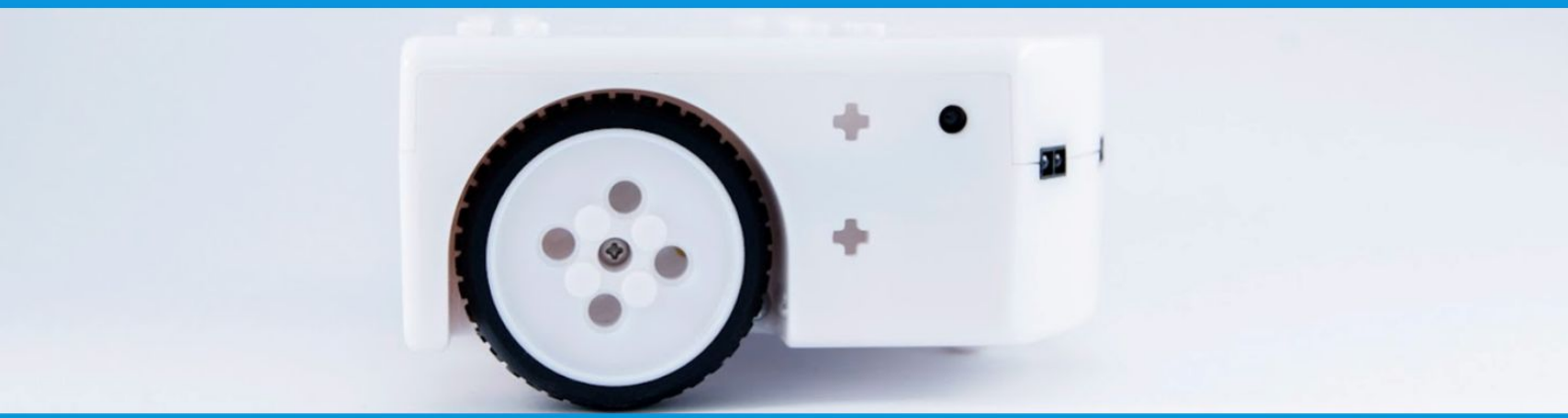
Action: move forward
Sensor: wall in front
Penalty ***in given!***

Connection - Settings

1. **Connect Thymio** with one of the following 2 methods:
    - either connect the robot to the computer using the USB cable
    - or connect the USB key and turn on the robot
2. **Launch** the **Thymio Suite software** and wait a few moments.
3. **Launch** the **Alphai software**.
4. In **Alphai**, select your **robot** that appears on the blue screen to **establish the connection**.

Step by step Table

Step by Step phase	Front sensors	Move backward to the left	Move forward to the left	Move forward	Move backward to the right	Move forward to the right	Reward value	Level value



www.thymio.org

thymio
by MOBSYA
Chemin du Closel 3, 1020 Renens - Switzerland
info@thymio.org