Human-centered Artificial Intelligence

2023/2024

Worksheet #2: Human-AI Cooperation in Multiagent Systems of Human Agents and Autonomous Artificial Intelligent Agents

Luís Macedo
Amílcar Cardoso, Hugo Oliveira

1.1 Topics

- Coordination vs Cooperation vs Collaboration
- Hybrid Human-Artificial Intelligence
- Autonomy, control, and automation
- Human in the loop
  - humans as data, information, and knowledge producers;
  - humans in the design, development and deployment of AI systems;
  - humans as recipients of AI systems output
- Human-in-the-Loop vs AI-in-the-Loop/Human-in-Command vs. Human-out-of-the Loop
- Human-AI Trustworthiness

1.2 Pre-class Readings

Look at the following resources:

- https://livebook.manning.com/book/human-in-the-loop-machine-learning/chapter-1/
- https://burrsettles.com/pub/settles.activelearning.pdf) (sections 1 and 2)

1.3 Theoretical Exercises

What's the difference between Cooperation and Collaboration? What's Hybrid Human-AI intelligence? What's Human in the Loop? What's the difference between Human in the Loop and AI-in-the-Loop (Human-in-Command or Machine in the Loop) vs. Human-out-of-the Loop?

1.4 Theoretical-Practical Exercises

Question 1.1 Consider the situations such as that of self-driving cars, Viking Sky, or Boeing 737 MAX (see https://www.aitrends.com/ai-insider/human-in-the-loop-vs-out-of-the-loop-in-ai-systems-the-case-of-ai-self-driving-

What's the involvement of humans in the machine processes in these situations?

Is it desirable? To what extent?

Isthi Human in the Loop (HITL)?

What are the advantages and disadvantages of HITL (as opposed to Human out of the Loop – HOOTL)?

Provide other examples of HITL and HOOTL situations.

Question 1.2

What is the dangerous of building an autonomous artificial agent that is able by himself to learn making decisions with no supervision and by interacting with the environment? What are the advantages and disadvantages?

As an example, take a look at these surprising results in intelligent behaviour:

https://www.youtube.com/watch?v=VMp6pq6_QjI

https://www.youtube.com/watch?v=GdTBqBnqhaQ

https://www.youtube.com/watch?v=Lu56xVlZ40M

Question 1.3 Assuming the collaboration between humans and AI agents is desirable, how to make it profitable? Which model to take? Human-in-the-Loop or Machine-in-the-Loop?

Question 1.4 Provide examples of Human-in-the-Loop and Machine-in-the-Loop situations?

Question 1.5 In which components of machines the collaboration between humans and the AI agent can occur and which way? Hint: take the typical ML pipeline, or the components of an AI agent (memory, sensors, effectors or actuators, learning, reasoning and decision-making, emotion, etc.).

Question 1.6 In order to be successfull, this Human-AI interaction requires especial features from the AI counterpart. It is not enough putting humans and AI agents collaborating, cooperating, negotiating. Is it important to make human-like AI?

Question 1.7 In HITL model, probably machines won't need always the help of humans. How to make machines decide when to ask the help of humans? How to make machines decide what to learn?

See the next example. How can we make the robot ask for help?

https://www.youtube.com/watch?v=NOLAwD4ZTW0

Question 1.8

Take a look at next video. Is it better to have a human-AI collaboration or should the robot scientist be able by himself to discover, to design experiments, to analyze data and generate hypothesis?

https://www.aber.ac.uk/en/cs/research/cb/projects/robotscientist/

https://www.youtube.com/watch?v=IY1sPV9e9H0

https://www.youtube.com/watch?v=2P5an-hJ-uU

Question 1.9

If we opt for a HOTL model in the case of the artificial scientist, is it possible to build it based only on the typical machine learning pipeline? Describe the traditional machine learning cycle and reflect on it. What kind of learning is there? What's the free will of the machine learning system?

Albert Einstein: "I have no special talent. I am only passionately curious".

What are the motivations behind learning in humans? Curiosity? What's curiosity?

In order to know more about human curiosity, take a look at:

https://www.youtube.com/watch?v=kKoNMYE9Bgs

Even monkeys have curiosity. So why not artificial agents?

To help, see:

https://www.youtube.com/watch?v=eof99XfoFfY
https://www.youtube.com/watch?v=7zpojhD4hpI

Question 1.10

Active learning is the mechanism that allows artificial agents to be curious. See examples of active learning in the following videos in which humans interact with the machine learning system. What are the advantages and disadvantages? Can the machine learning system be autonomous and select what to learn?

Take a look at the next videos:
https://www.youtube.com/watch?v=D8Q8Udh7CMg
https://www.youtube.com/watch?v=B9jwNWiEd_M
https://www.youtube.com/watch?v=chukkEeGrLM
https://www.youtube.com/watch?v=NOLAwD4ZTW0
https://www.youtube.com/watch?v=S1wXI35GG7k

Question 1.11

What's the difference between passive and active learning?

Question 1.12 The question is how to implement active learning strategies on agents.

Consider the problem of respiratory diseases. Suppose there is a data set D with a description of symptoms associated with a disease. Suppose also that there are three new cases, x1, x2 and x3, that can be used to update the data set of the AI agent. Now, it is important to know whether someone has Covid-19 or not. Assume that, based on a previous data set, the agent computes the following probabilities for the labels of Covid-19 for those cases:

Instance	P(Covid=T)	P(Covid =F)
X1	0.85	0.15
X2	0.60	0.40
X3	0.50	0.50

a) Which instance should be selected for labeling first by an oracle, i.e., for which instance should the agent ask help to a human expert, if we use the Least Confident, the Smallest Margin and the Entropy-based approaches?

b) Consider now that Covid-19 can have three labels (High, Medium, Low). Consider also that the probability distributions for the three

instances are now the ones shown in the table below. Which instance should now be selected for labeling first by an oracle if we use the Least Confident, the Smallest Margin and the Entropy-based approaches?

Instance	P(Covid=High)	P(Covid =Medium)	P(Covid =Low)
X1	0.70	0.25	0.05
X2	0.32	0.32	0.36
X3	0.95	0.05	0

Question 1.13 Consider now the situations in the next videos. They represent a different form of interaction/collaboration between an AI agent and humans. What are the differences to the previous videos? What information is given by humans to the robots?

http://www.scholarpedia.org/article/Robot_learning_by_demonstration

https://www.youtube.com/watch?v=ta9SkcJdSBA

1.5 Post-class Readings

Read the following texts (available in the course's repository in UCStudent and/or in the Internet):

- https://freecontent.manning.com/10-ways-that-human-in-the-loop-machine-lear
- https://link.springer.com/article/10.1007/s10462-022-10246-w
- https://medium.com/vsinghbisen/what-is-human-in-the-loop-machine-learning-w
- https://www.aitrends.com/ai-insider/human-in-the-loop-vs-out-of-the-loop-in-ai-systems-the-case-of-ai-self-dri
- Settles [2009]
- https://burrsettles.com/pub/settles.activelearning.pdf)
- https://arxiv.org/pdf/1910.05789.pdf
- https://www.dfki.de/en/web/research/research-departments/interactive-machine-learning/

Bibliography

Burr Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009. URL http://axon.cs.byu.edu/~martinez/classes/778/Papers/settles.activelearning.pdf.