Human-centered Artificial Intelligence

2024/2025

Worksheet #3:
Human in the loop machine learning (part I):

Luís Macedo
Amílcar Cardoso, Hugo Oliveira

1.1 Topics

- Sampling training data for human annotation

  - Curious agents and Active Learning
  - Active Learning Scenarios (stream-based, pool-based, query synthesis)
  - Query strategies (Uncertainty sampling; Diversity sampling; Hybrid sampling; Transfer learning-based sampling)

- Overview of annotation

- AI agents learning to defer/delegate decision-making

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

Question 1.1 What are the limitations of passive machine learning?

What are the advantages of curious agents?

What is active learning? What is the relation to HitL?

Distinguish the different Active Learning Scenarios (stream-based, pool-based, query synthesis).

Distinguish the different Query Strategies (Uncertainty sampling; Diversity sampling; Hybrid sampling; Transfer learning-based sampling) Active Learners can employ.

How can AI agents learn to defer/delegate tasks?

1.4 Theoretical-Practical Exercises

Question 1.2

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 1.3

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=NOLAwD4ZTWO
https://www.youtube.com/watch?v=S1wXI35GG7k

Question 1.4 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

Least Confident:
  X1 = 1 - (maior valor) = 1 - 0.85 = 0.15
  X2 = 1 - 0.6 = 0.4
  X3 = 1 - 0.5 = 0.5
    -> We would select X3, is the highest

Smallest Margin:
  X1 = 0.85 - 0.15 = 0.7
  X2 = 0.6 - 0.4 = 0.2
  X3 = 0.5 - 0.5 = 0
    -> We would select X3, is the lowest

Entropy:
  X1 = - 0.85*log(0.85) - 0.15*log(0.15)
    -> After doing the maths, we would
       also choose X3

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 1.5 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.