1 Overview

The aim of this project is to build a Machine Learning (ML) system to analyze the IPARC dataset. This is designed to test the ability to frame hypothesis and to design ML based experiments to verify or reject the hypothesis. To get you started, we provide a couple of examples on how to proceed. But you are expected to come up with your own hypothesis.

2 Background

Artificial Intelligence (AI) is the one of the broad aims of machine learning systems. Even with the breakthroughs such as ChatGPT, AI seems to be quite difficult. A part of this can be attributed to the lack of conceptual clarity on the meaning of "intelligence". In the article [Cho19] François Chollet discusses the definition of intelligence and introduces the `Abstraction and Reasoning Corpus` to benchmark the progress towards AI. However, not much progress could be made as can be seen from the leader-board at https://www.kaggle.com/competitions/abstraction-and-reasoning-challenge/overview.

A much more simpler IPARC dataset (https://github.com/ac20/IPARC_ChallengeV2) can also be used. This resembles the ARC dataset but is much more easier to analyze since the data generation process is known. Please go through the presentation and the article (links available in the description of the above repo) for details.

3 Project Statement

In this project you are expected to frame an hypothesis about intelligence and current ML algorithms and test it using either the original ARC dataset or the simpler IPARC dataset. You are required to form a group of 3-4 members. You should submit a report (not exceeding 5 pages [1]) by April 15th 2023 with the following sections

1. Problem Statement
2. Methodology
3. Experimental Results and Validation
4. Conclusion and Future Work

Also, you or one of your group members may be asked to give a presentation of around 10 minutes at the end of the semester.

To get you started we provide a couple of ideas and what is expected in each section below.

3.1 Problem Statement

What constitutes a good problem statement? - A good problem statement should explicitly state what the goal of the project is. And moreover it should also include a measure of the extent to which the problem is solved. Usually a good problem statement has a hypothesis which can be shown to be either true or false.

Example 1. An IPARC task consists of a sequence of functions like output $= \phi_k \circ \phi_{k-1} \circ \cdots \circ \phi_1(\text{input})$, where each of the ϕ_i come from a pre-defined set of functions. A good question would be - Given the input and output can we train a linear regression or a decision tree model to check if a given ϕ_i is in the sequence or nor?

[1] Reports exceeding 5 pages will be penalized

Example 2. A fundamental problem with artificial intelligence (coming from ARC) can be watered down to the following problem - Say you have observations $\{x_i, y_i\}$, where y_i are the labels. Task-1 requires you to predict y_i when the input is x_i. And Task-2 requires you to predict $-1 * y_i$ when the input is x_i. Now given a input x_i (you have no knowledge if it's task-1 or task-2) does there exist a classifier which predicts the output?

3.2 Methodology

For problems which include a well framed hypothesis, methodology involves framing an experiment. A good experiment is the one which either proves or disproves your hypothesis. In other words, it should not leave you uncertain with regards to the hypothesis.

Example 1. For instance in the example 1 above, if we replace the question with - Given the input and output can we *find a model* to check if a given ϕ_i is in the sequence or nor?, then you can spend your entire life searching through the models and not know if this is possible or not! However, because you have decided to use linear functions or decision trees it is much more easy to do the experiment.

-

For theoretically oriented problem statements, a good methodology involves seeing how much of the current known theory is consistent with the problem statement and "twisting" the question to give positive answers.

Example 2. It is clear from Bayes theory that there cannot exist a classifier which gives both output y_i and $-y_i$ given x_i. That is a system, no matter how intelligent, cannot identify if it is in task-1 or task-2 in this scenario. What is missing? - There needs to be some form of information apart from x_i to identify whether it is in task-1 or task-2. Suppose we add a small error $\epsilon > 0$ to the inputs of task-1, would it then be possible? How hard would it be to learn such a function? ('Hardness' in ML is measured in terms of number of samples required to learn the function)

As you might have gathered this is an open-ended task. As to how much is expected in this project is discussed in evaluation section below.

3.3 Experimental Results and Validation

It is rarely the case that we trust a piece of information coming from a single source. All information should be validated from different perspectives to ensure correctness.

If your problem statement involves an hypothesis, and you have shown it to be TRUE/FALSE you should discuss how well it correlates with current knowledge, and/or identify what aspects made this hypothesis TRUE/FALSE.

Example 1. In example 1. (say) you found out that linear classifiers give 20% accuracy and decision trees give 80% accuracy. What does it mean? Can you identify where the mistakes of the model were made? (a.k.a Error Analysis)

For theoretically oriented problem statements, experiments are usually on simple datasets (such as IPARC) whose behaviour is well understood. This allows inference on the theoretical aspects.

Example 2. Say you have decided to create a dataset (using IPARC source code) such that if the pixel on the top-right corner is 1 then you use one operator ϕ_1 else you use another operator ϕ_2. You can check how many examples would it take for (a) Linear classifier and (b) Decision Tree to identify this rule (top-right corner is 1). What if you change the rule slightly to top-right corner is [[1,1],[1,1]] or otherwise?

3.4 Conclusion and Future Work

You should then summarize your work in simple words and identify precisely what you have proved. More importantly, you should also state what you have not proved and should be considered for future work

Example 1. Say you have achieved a 80% accuracy using decision trees. It probably means that for the IPARC dataset you can fairly predict if the operator is within the sequence or not. However, in the IPARC dataset we only have 8 operators. It does not show what would happen if the number of operators is increased to 2^8 instead! This is to be checked in future work.

Example 2. Suppose you have found out after extensive experiments that one cannot distinguish between task-1 and task-2 for small $\epsilon > 0$ with linear or decision tree classifiers. The issue can be traced to the local smoothness which is required by these class of functions. But there might exist another class of functions which are capable of it.

4 Evaluation of the projects

As stated above, these projects are expected to be done in groups of 3-4. The evaluation would reflect the standard peer-review procedure to evaluate scientific literature. With an exception that the peers shall be the instructors/TAs. In the sense that,

- If we can identify holes in your arguments easily, they would be penalized. For instance, if you have used the same set for both train and validation, it would be heavily penalized.
- Originality and Creativity would be rewarded. If the questions and results you generate are novel and surprising then it would be highly rewarded.
- You are encouraged to go through the video at https://youtu.be/SPVWSG7-i_E?t=1742 and slides at https://drive.google.com/file/d/15hPTA64h31ShaoybLWeU3moZan7zVbr_/view?usp=sharing (made for the top conference on ML) to get an idea of the peer-review procedure.

5 Final words

This project constitutes of 10% of your grade. While the remaining 90% is designed to evaluate your technical ability, the aim of this project is to test some important soft-skills such as (a) Critical Thinking and (b) Communication of ideas. While the actual results you get from this project may/may-not be useful in your career progression, we can assure you that the soft-skills you acquire would be a great deal beneficial. Hope you enjoy this excercise!

References

[Cho19] François Chollet. "On the Measure of Intelligence". In: *CoRR* abs/1911.01547 (2019). arXiv: 1911.01547. URL: http://arxiv.org/abs/1911.01547.