Evento	Salão UFRGS 2019: SIC - XXXI SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2019
Local	Campus do Vale - UFRGS
Título	ON THE REASONING CAPABILITIES OF LANGUAGE MODELS
Autor	RAFAEL BALDASSO AUDIBERT
Orientador	LUIS DA CUNHA LAMB

INSTITUTO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Project ON THE REASONING CAPABILITIES OF LANGUAGE MODELS

Author: RAFAEL BALDASSO AUDIBERT
Supervisor: Dr. LUÍS DA CUNHA LAMB

Introduction:

The research theme has been established as trying to understand what (and perhaps, how) a powerful neural network based on deep learning can learn, given the correct training dataset and the necessary number of parameters (or neurons).

Recently, OpenAI "*trained a large-scale unsupervised language model which generates coherent paragraphs of text" (RADFORD et al, 2019)* called GPT-2. The machine learning and natural language processing communities have been positively impressed with the model's capability of sustaining long-term consistency throughout sentences. On the other hand, some researchers expressed concerns that GPT-2 is merely memorizing an extensive dataset of texts instead of effectively reasoning about what it says.

Currently, OpenAI released only 2 models for the general public with 117 and 345 million parameters, due to their concerns on malicious applications of the technology. Our project, then, attempts to go further and investigate whether the model can reason about different things than what it was created for.

Method:

We developed some tasks to be solved by the model, described below. Each task is accompanied by the set of abilities we deem are required to solve it: (i) First Order Logic (FOL) Solver - Represent semantical proof structure as a graph similar to an Abstract Syntax Tree; (ii) Explainable linear equations system solver - Learn to compute matrix properties (such as determinant), generating a step-by-step guide towards the solution; (iii) Mathematical Reading and Solving - Learn to evaluate texts as symbolic expressions and manipulate them.

As stated before, we will use the already released GPT-2 trained models to leverage the semantics knowledge, such that we just need to learn the specifics for each task. As a baseline, we will train the tasks without the pre-trained parameters, to see if the semantics learned by the basic model help in the writing ability needed on the second and third task.

Contribution:

This project's contribution lies in the field of understanding what neural networks can infer from text-based information, as well as creating three datasets (and the code that generates it, in Ruby), one for each of the tasks proposed above, as there are no publicly available big datasets on these areas, especially for the FOL Solver task.

Results:

The project is currently in the model training phase, but preliminary results suggest that the models can solve the tasks proposed. The final results will be presented at SIC 2019.

Bibliography:

RADFORD et al., **Language Models are Unsupervised Multitask Learners**, 2019, preprint: available at https://github.com/openai/gpt-2

JUNE, 2019