Paper reports – Mohammad Osoolian

Paper Title: HellaSwag: Can a Machine Really Finish Your Sentence?

Paper Link: [1905.07830] HellaSwag: Can a Machine Really Finish Your Sentence? (arxiv.org)

Submit Date: 19 May 2019

What is paper about: This paper introduces HellaSwag dataset for NLI taks which as an update for SWAG dataset.

Abstract:

HellaSwag introduces a new challenge dataset for evaluating commonsense natural language inference (NLI). The dataset, called HellaSwag, is designed to be difficult for even state-of-the-art NLI models, while still being easy for humans.

Background:

- NLI
- SWAG
- AF
- BERT

Challenge:

Existing datasets for NLI are not always challenging enough for state-of-the-art models, which can lead to overfitting and inaccurate results. HellaSwag is a more challenging dataset that can help to ensure that NLI models are truly capable of understanding commonsense reasoning.

New Ideas:

The prompts in HellaSwag are longer and more complex than those in previous datasets. This makes it more difficult for models to understand the context of the prompt and to generate a relevant response.

The endings in HellaSwag are designed to be adversarial, meaning that they are deliberately designed to be difficult for models to distinguish from human-written endings. This makes it more challenging for models to learn to trust their own predictions.

Results:

The results of HellaSwag show that it is a significantly more difficult dataset for NLI models than existing datasets. For example, the state-of-the-art model, BERT, achieves only 48% accuracy on HellaSwag, compared to 90% on other datasets.

My Idea for the challenge:

Make model to explain the reasons of chooses to find what the Annotation Artifacts are in the dataset and prevent them.

My Idea to improve this article:

Working on new model or an improvement of BERT for the new challenging dataset: HellaSwag