Paper reports – Mohammad Osoolian

Paper Title: Annotation Artifacts in Natural Language Inference Data

Paper Link: [1808.05326v1] SWAG: A Large-Scale Adversarial Dataset for Grounded Commonsense Inference (arxiv.org)

Submit Date: 16 Aug 2018

What is paper about: This paper introduces a new dataset called SWAG that bring a new challenge for state-of-art NLI models.

Abstract:

The SWAG dataset is constructed by first collecting pairs of temporally adjacent video captions, each with a context and a follow-up event that we know is physically possible. The authors then use a state-of-the-art language model fine-tuned on this data to massively oversample a diverse set of possible negative sentence endings (or counterfactuals). This results in a dataset that is large-scale and challenging, requiring models to reason about the physical world and commonsense knowledge.

Background:

- NLI
- SNLI
- Annotation Artifacts in NLI datasets

Challenge:

NLP models have reached the human accuracy for NLI tasks on famous datasets like SNLI and MNLI. However, models are still not powerful enough to get used in real tasks. Therefore, there is need for a more complex dataset for NLI to be bigger and more variant.

New Ideas:

SWAG solves this challenge by creating a dataset of 113k multiple choice questions from variety of situations. SWAG uses Adversarial Filtering to de-bias the dataset. This means that SWAG is less likely to be biased by human annotators, which can be a problem with other datasets.

Results:

The authors evaluated SWAG on a variety of models, including BERT, RoBERTa, and XLNet. They found that SWAG is a more challenging dataset than existing benchmarks, such as SNLI and MNLI. They also found that SWAG can be used to improve the performance of models on these benchmarks.

The authors also evaluated the effectiveness of Adversarial filtering. They found that AF can significantly reduce the bias in datasets for grounded commonsense inference. This makes SWAG a more reliable resource for research on this task.

My Idea for the challenge:

using more workers for generating dataset in a way that every worker doesn't write more than 5 samples of dataset could prevent Annotation Artifacts in dataset.

My Idea to improve this article:

Focusing on creating de-baiased datasets for others tasks in NLP using the Adversarial Filtering.