Paper reports – Mohammad Osoolian

Paper Title: Annotation Artifacts in Natural Language Inference Data

Paper Link: [1803.02324] Annotation Artifacts in Natural Language Inference Data (arxiv.org)

Submit Date: 6 Mar 2018

What is paper about: This paper discusses bout the artifacts in human generated NLI datasets that make the model learn the correct hypothesis without seeing the premise

Abstract:

The article discusses how large-scale datasets for natural language inference (NLI) are created by having crowd workers generate sentences based on given premises. However, it's shown that clues in the data enable labels (entailment, contradiction, neutral) to be predicted solely from the generated sentences without considering the premises. This undermines the effectiveness of NLI models, suggesting that their success has been overestimated. Specific linguistic patterns like negation and vagueness are linked to certain types of inference. This highlights the difficulty of the NLI task as an ongoing challenge

Background:

- NLI
- SNLI
- MNLI
- Image manipulation techniques

Challenge:

certain linguistic cues and patterns in the generated hypotheses of natural language inference datasets can reveal the correct inference label (entailment, contradiction, neutral) without needing to consider the original premises.

New Ideas:

Train a model to predict the hypothesis labels without seeing the premise. The correct predictions samples are classified as “easy” samples and wrong predictions are classified as “hard” samples. The test state-of-art NLP models on hard samples to check the real accuracy of models.

Results:

Accuracy of DAM, ESIM and DIIN models was tested on SNLI and MNLI and in all cased the real accuracy of models which is the accuracy tested on “hard” samples was lower than the total accuracy of models. This shows that the models have learnt to predict hypothesis labels without attention to the premise.

My Idea for the challenge:

Working on approaches to create a NLI dataset by machines and random actions. Datasets created by human may be effected by habits of workers but with randomness of machines the problem can be solved.

My Idea to improve this article:

Finding the repeated paradigms in “easy” samples that makes model able to predict hypothesis without seeing the premise.