Harnessing Privileged Information for Hyperbole Detection

Rhys Biddle[1,2], Maciek Rybinski[2], Qian Li[1], Cécile Paris[2], Guandong Xu[1]
[1] Advanced Analytics Institute, University of Technology Sydney, Australia
[2] CSIRO Data61, Sydney, Australia

INTRODUCTION

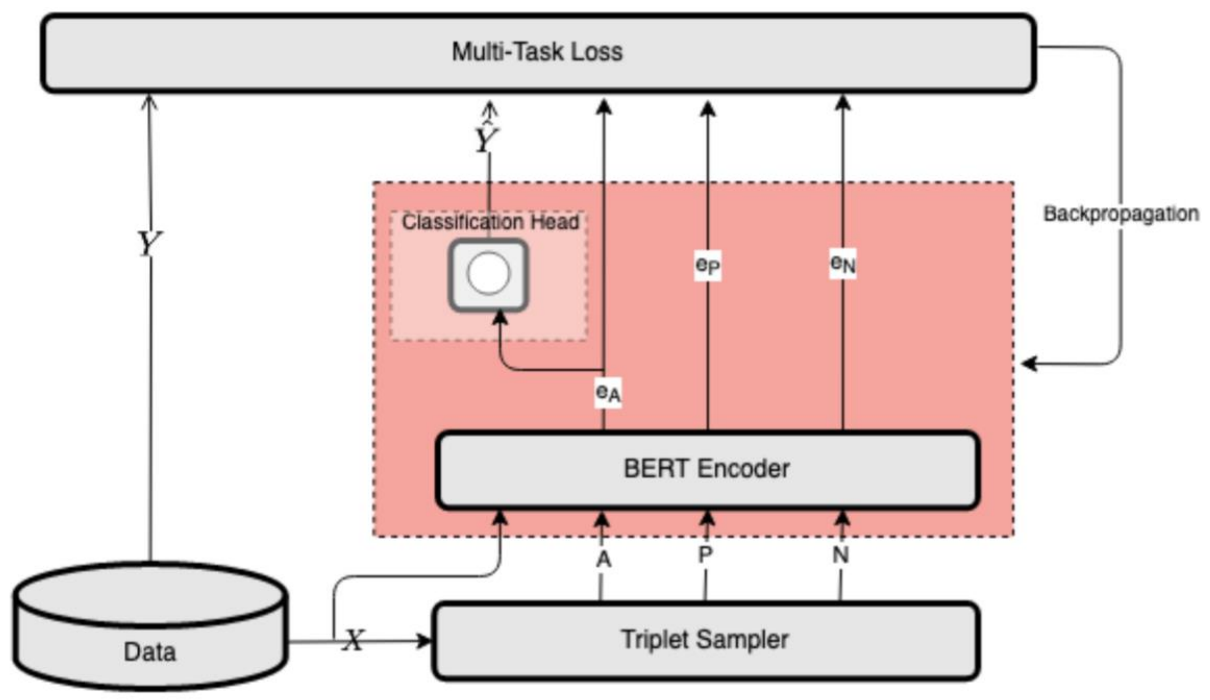
We propose a model that combines pre-trained language models with privileged information for the task of hyperbole detection. Experiments show that our model improves upon baseline models on an existing hyperbole detection dataset. Further, we discover that our experiments highlight annotation artifacts introduced through the process of literal paraphrasing of hyperbole

AIM

The detection of hyperbolic text in short sequences is a challenging task and we propose to use literal interpretations to address this challenge. A small dataset of probing examples is designed to gain further insight into the hyperbole detection task.

METHODS

BERT+PI is based on a multi-task sequence classification architecture. The model is optimised via combination of cross entropy loss and a triplet loss.



$$\hat{y}_i = \sigma(e_i^a \mathbf{W} + b),$$

$$\mathcal{L} = \mathcal{L}_c + \lambda \mathcal{L}_t$$

$$\mathcal{L}_c = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right]$$

$$\mathcal{L}_t = \frac{1}{Ns}\sum_{i=1}^{N}\sum_{j=1}^{s}\left[\max(D(e_i^a, e_{ij}^p) - D(e_i^a, e_{ij}^n) + m, 0)\right]$$

Figure 1: BERT+PI

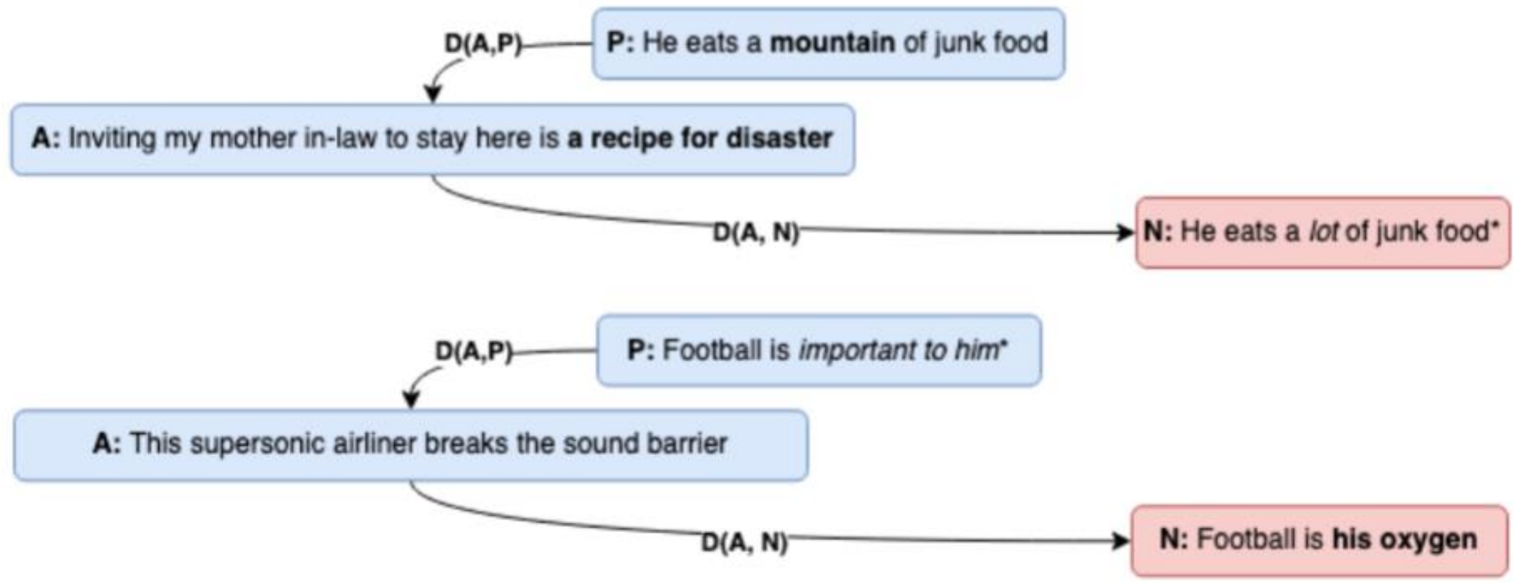


Figure 2: Triplet Sampling

RESULTS

BERT+PI showed significant performance improvements over existing models for hyperbole detection and base **BERT** on an existing hyperbole detection benchmark.

Model	F1	Precision	Recall
LR+QQ	0.710(-)	0.679(-)	0.745(-)
NB+QQ	0.693(-)	0.689(-)	0.696(-)
BERT	0.709(.064)	0.711(.077)	0.735(.177)
BERT+QQ	0.671(.086)	0.650(.147)	0.765(.246)
BERT+PI	**0.781(.012)**	**0.754(.053)**	**0.814(.039)**

Table 1: Hyperbole Detection Results

However, probing experiments revealed that **BERT+PI** excelled at detecting one type of hyperbole (i.e., Extreme Case Formulations). Results on quantitative and qualitative were poor, similar to other baselines.

DISCUSSION

Explanations indicated that **BERT+PI** was able to differentiate between hyperbolic and non-hyperbolic expression of hyperbole-prone words in some instances.

BERT		BERT+PI	
LIME Word Weightings	P(*h*)	LIME Word Weightings	P(*h*)
Search engines are brainless entities.	.66	Search engines are brainless entities.	.18
Me, the wife of that boorish, brainless man.	.78	Me, the wife of that boorish, brainless man.	.74
This policy will plunge the country into a chaos.	.20	This policy will plunge the country into a chaos.	.79
Every flavor is dynamite.	.35	Every flavor is dynamite.	.96

Figure 3: Understanding Context

However, explanations also indicated errors from exploiting annotation artifacts.

BERT+PI	
LIME Word Weightings	P(*h*)
her brain is as small as a quarter	.86
Her hair is as thin as silk	.84
my heart is as heavy as the world	.73
his mouth is as big as a barn	.87
His beard is as thick as his mustache	.87
that bag is as heavy as a suitcase	.72
Her sister is as tall as her mother	.86
their hair is as long as a finger	.74

Figure 4: Exploiting Annotation Artifact

CONCLUSIONS

The incorporation of privileged information into a multi-task classification resulted in significant improvements over existing hyperbole detection models.

However, probing experiments indicated the incorporation of literal paraphrases as privileged information was helpful in only some hyperbolic expressions. Particularly, when the hyperbolicity was contained within a few words. More complex hyperbolic expressions were poorly handled by all models and remain a focus of future research