



Adverb Is the Key: Simple Text Data Augmentation with Adverb Deletion

Juhwan Choi YoungBin Kim
gold5230@cau.ac.kr ybkim85@cau.ac.kr

Chung-Ang University



Preliminary & Motivation

- In the field of text data augmentation, rule-based methods such as **EDA**¹ are widely used
- However, such methods may harm the semantics of original sentence
- We propose a simple rule-based technique that preserves the semantics through the **deletion of adverbs**

1. Wei and Zou, *EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks*, EMNLP 2019.

Operation	Sentence
None	A sad, superior human comedy played out on the back roads of life.
SR	A <i>lamentable</i> , superior human comedy played out on the <i>backward</i> road of life.
RI	A sad, superior human comedy played out on <i>funniness</i> the back roads of life.
RS	A sad, superior human comedy played out on <i>roads</i> back <i>the</i> of life.
RD	A sad, superior human out on the roads of life.

Method

- We analyze the dataset through POS tagger and delete adverbs from each sentence
- Sentences without adverbs are jointly used for training the model
- $\hat{x} = x \setminus \{W_{adv}\}$, where $\{W_{adv}\} = POS_TAGGER(x, ADV)$

Original	EDA	Ours
The film is strictly routine.	The is strictly film routine.	The film is routine.
The santa clause 2 proves itself a more streamlined and thought out encounter than the original could ever have hoped to be.	The santa clause 2 itself a streamlined and original thought out encounter than the could have to be.	The santa clause 2 proves itself a streamlined and thought out encounter than the original could have hoped to be.
This is a very ambitious project fora fairly inexperienced filmmaker, but good actors, good poetry and good music help sustain it.	This is a very ambitious for a fairly inexperienced actors, but filmmaker good, good poetryand good music sustain it.	This is a ambitious project for a inexperienced filmmaker, but good actors, good poetry and good music help sustain it.
Perhaps the best sports movie i've ever seen.	Best perhaps movie seen.	the best sports movie i've seen.

Experiment & Conclusion

- We trained BERT² model for various NLI datasets with different data augmentation methods
- We found that our method is especially effective for NLI tasks as it tends to preserve the semantics of the sentence
- Additionally, the usage of curriculum data augmentation offered additional performance gain

BERT	SST2	SST5	CoLA	TREC	RTE	MNLI-M	MNLI-MM	QNLI
No Aug	88.74	50.58	79.24	95.08	64.72	75.35	77.17	85.81
EDA	89.18	50.36	76.80	95.27	62.43	74.51	75.85	83.17
AEDA	89.41	50.63	79.45	95.39	65.92	74.60	77.35	86.55
softEDA	89.24	50.89	76.32	95.86	65.41	74.06	75.93	84.97
Ours	89.73	51.25	80.28	96.48	65.58	76.11	77.68	87.07
Ours w/ Curr.	89.52	52.37	83.15	96.95	68.72	77.97	78.14	87.21

2. Devlin et al., *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, NAACL 2019.