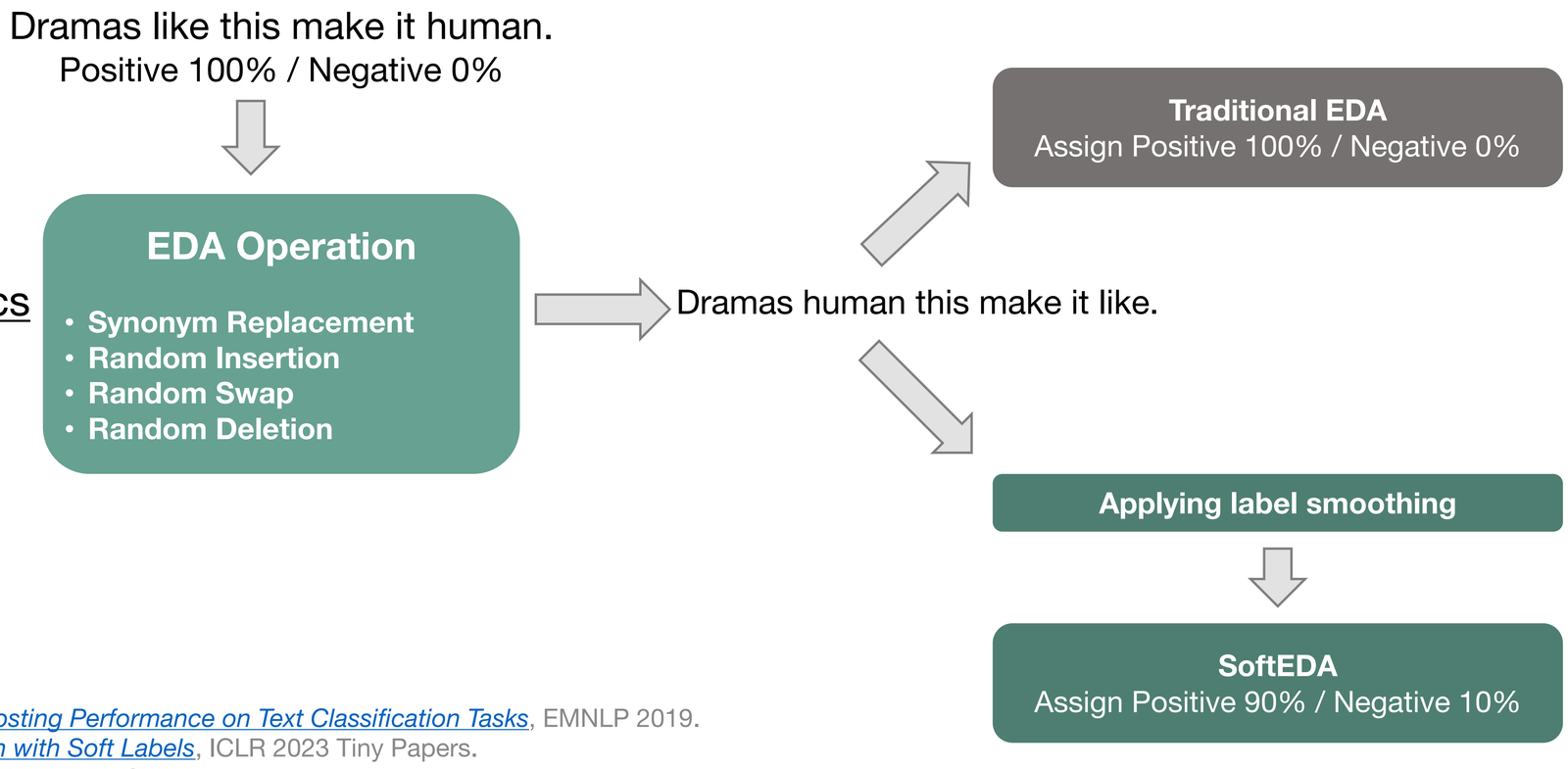
AutoAugment Is What You Need: Enhancing Rule-based Augmentation Methods in Low-resource Regimes

Juhwan Choi gold5230@cau.ac.kr
Kyohoon Jin fhzh123@cau.ac.kr
Junho Lee jhjo32@cau.ac.kr
Sangmin Song s2022120859@cau.ac.kr
YoungBin Kim ybkim85@cau.ac.kr
Chung-Ang University

Preliminaries: EDA & SoftEDA

- In the field of text data augmentation, **EDA**[1] is widely used technique
- However, EDA may <u>harm the semantics</u> of original sentence
- Recently, a novel technique called **SoftEDA**[2] suggested to compensate the damage with label smoothing[3]



1. Wei and Zou, *EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks*, EMNLP 2019.
2. Choi et al., *SoftEDA: Rethinking Rule-Based Data Augmentation with Soft Labels*, ICLR 2023 Tiny Papers.
3. Szegedy et al., *Rethinking the Inception Architecture for Computer Vision*, CVPR 2016

Motivation

- SoftEDA manually assigned a factor for label smoothing
- Despite of the strength of SoftEDA, this drawback **decreases the real-world applicability**
- We apply AutoAugment[4] to solve this issue and **automatically find optimal value**
- Furthermore, we aim to enhance the performance of cutting-edge level of PLM with the proposed method, <u>not just BERT</u>

4. Cubuk et al., *AutoAugment: Learning Augmentation Strategies from Data*, CVPR 2019

Method

- Following previous work[5], we design an augmentation policy including:
  - Number of augmented data, amount of label smoothing factor, …
- We optimize the policy based on sequential model-based global optimization[6]
  - This approach offers **better solution than inefficient grid search**

5. Ren et al., *Text AutoAugment: Learning Compositional Augmentation Policy for Text Classification*, EMNLP 2021
6. Bergstra et al., *Algorithms for Hyper-Parameter Optimization*, NeurIPS 2011

Experiment and Result

- We conducted experiment with BERT and DeBERTaV3 in <u>low-resource scenario</u> (100 and 500 original data)
- In DeBERTaV3, <u>baselines faced performance degradation</u>
- Whereas, our method consistently improved the model
- We found that **rule-based augmentation is still beneficial** with carefully designed policies

	SST2	SST5	CoLA	SUBJ	TREC	MR	CR	PC
DeBERTaV3 w/o Aug	$88.36_{0.36}$	$35.95_{1.69}$	$72.62_{4.24}$	$92.23_{0.24}$	$80.19_{3.23}$	$82.84_{0.39}$	$85.61_{1.20}$	$91.22_{0.43}$
	$92.59_{0.73}$	$48.77_{1.52}$	$82.21_{0.82}$	$94.66_{0.22}$	$94.06_{0.43}$	$86.22_{0.37}$	$91.40_{0.36}$	$91.85_{0.26}$
w/ EDA	$86.61_{0.70}$	$37.64_{1.23}$	$74.83_{1.10}$	$92.85_{0.48}$	$83.65_{1.84}$	$83.18_{\mathbf{0.32}}$	$84.86_{0.73}$	$90.51_{0.47}$
	$93.25_{0.55}$	$49.04_{0.78}$	$79.24_{0.66}$	$94.81_{0.53}$	$94.33_{0.99}$	$86.71_{0.65}$	$91.24_{0.39}$	$92.3_{0.15}$
w/ AEDA	$88.44_{0.80}$	$36.87_{2.88}$	$79.29_{0.65}$	$92.81_{0.47}$	$84.17_{0.79}$	$82.87_{0.75}$	$85.76_{1.37}$	$90.61_{0.49}$
	$92.54_{0.78}$	$49.16_{0.83}$	$82.78_{\mathbf{0.40}}$	$94.92_{0.58}$	$94.45_{0.80}$	$85.77_{1.63}$	$91.09_{0.49}$	$92.29_{\mathbf{0.11}}$
w/ softEDA	$88.94_{1.03}$	$38.37_{1.65}$	$79.40_{1.51}$	$92.90_{1.08}$	$84.58_{1.29}$	$83.50_{0.65}$	$86.33_{1.65}$	$91.28_{0.82}$
	$93.12_{1.05}$	$50.34_{1.44}$	$78.97_{1.16}$	$94.77_{0.21}$	$94.71_{0.69}$	$87.02_{0.50}$	$91.81_{0.76}$	$92.16_{0.20}$
w/ Ours	$\mathbf{91.38}_{0.32}$	$\mathbf{42.92}_{0.52}$	$\mathbf{82.56}_{0.51}$	$\mathbf{94.47}_{0.26}$	$\mathbf{87.70}_{0.90}$	$\mathbf{85.31}_{0.79}$	$\mathbf{89.95}_{\mathbf{0.51}}$	$\mathbf{92.32}_{\mathbf{0.19}}$
	$\mathbf{93.94}_{\mathbf{0.30}}$	$\mathbf{52.77}_{\mathbf{0.62}}$	$\mathbf{84.32}_{0.49}$	$\mathbf{95.29}_{0.31}$	$\mathbf{94.92}_{0.62}$	$\mathbf{87.96}_{\mathbf{0.17}}$	$\mathbf{92.46}_{\mathbf{0.18}}$	$\mathbf{92.72}_{0.40}$
w/ Ours w/o LS	$90.47_{\mathbf{0.26}}$	$42.44_{\mathbf{0.49}}$	$82.10_{\mathbf{0.43}}$	$94.22_{\mathbf{0.15}}$	$86.57_{\mathbf{0.61}}$	$85.07_{0.58}$	$89.47_{0.67}$	$92.22_{0.21}$
	$93.40_{0.58}$	$52.54_{0.66}$	$83.67_{0.86}$	$95.15_{\mathbf{0.12}}$	$94.92_{\mathbf{0.18}}$	$87.41_{0.37}$	$92.28_{0.27}$	$92.49_{0.33}$