SoftEDA: Rethinking Rule-Based Data Augmentation with Soft Labels

• Juhwan Choi, Kyohoon Jin, Junho Lee, Sangmin Song, Youngbin Kim*

Intelligent Information Processing Lab.

IIPL

Juhwan Choi

gold5230@cau.ac.kr

Intro: Importance of Data

Deep Learning is data-hungry strategy

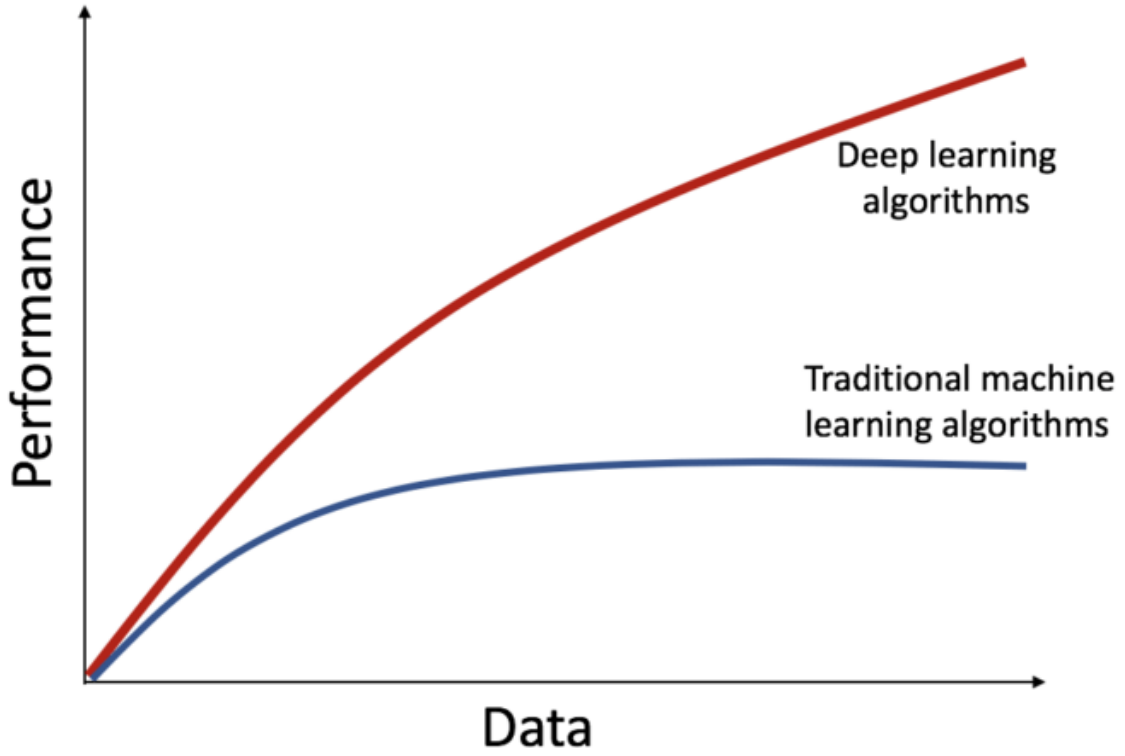


Image from https://abyssal.eu/were-data-hungry/

Intro: Importance of Data

Insufficient data may lead to overfitting

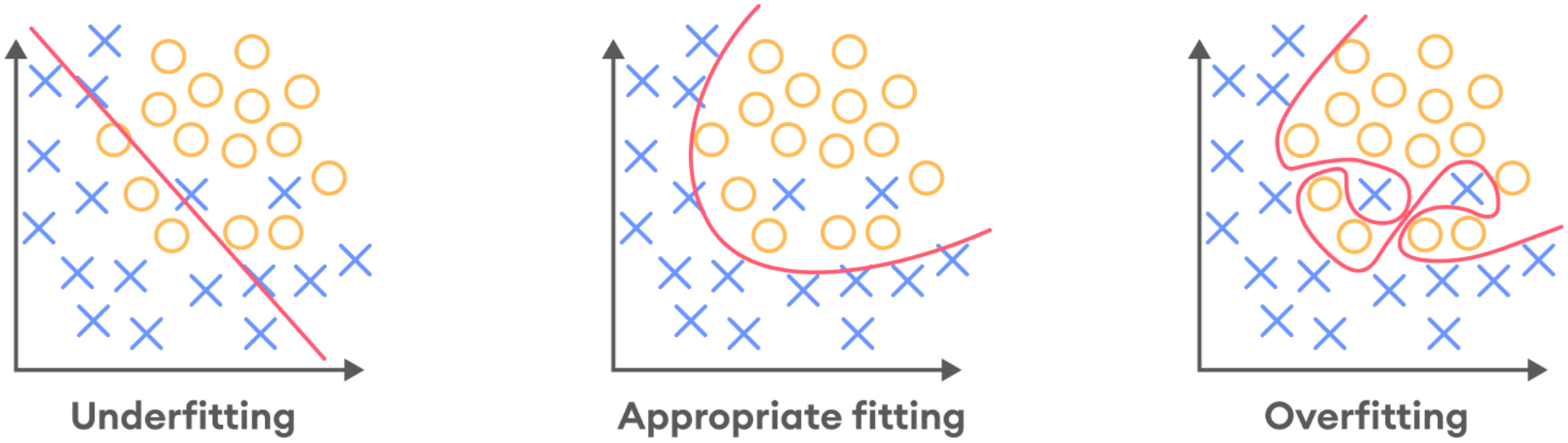


Image from https://medium.com/@hassankashi.official/deep-learning-convolutional-neural-networks-cnn-recurrent-neural-networks-rnn-lstm-long-7ba53032b46c

Intro: Data Augmentation

A common approach to mitigate overfitting

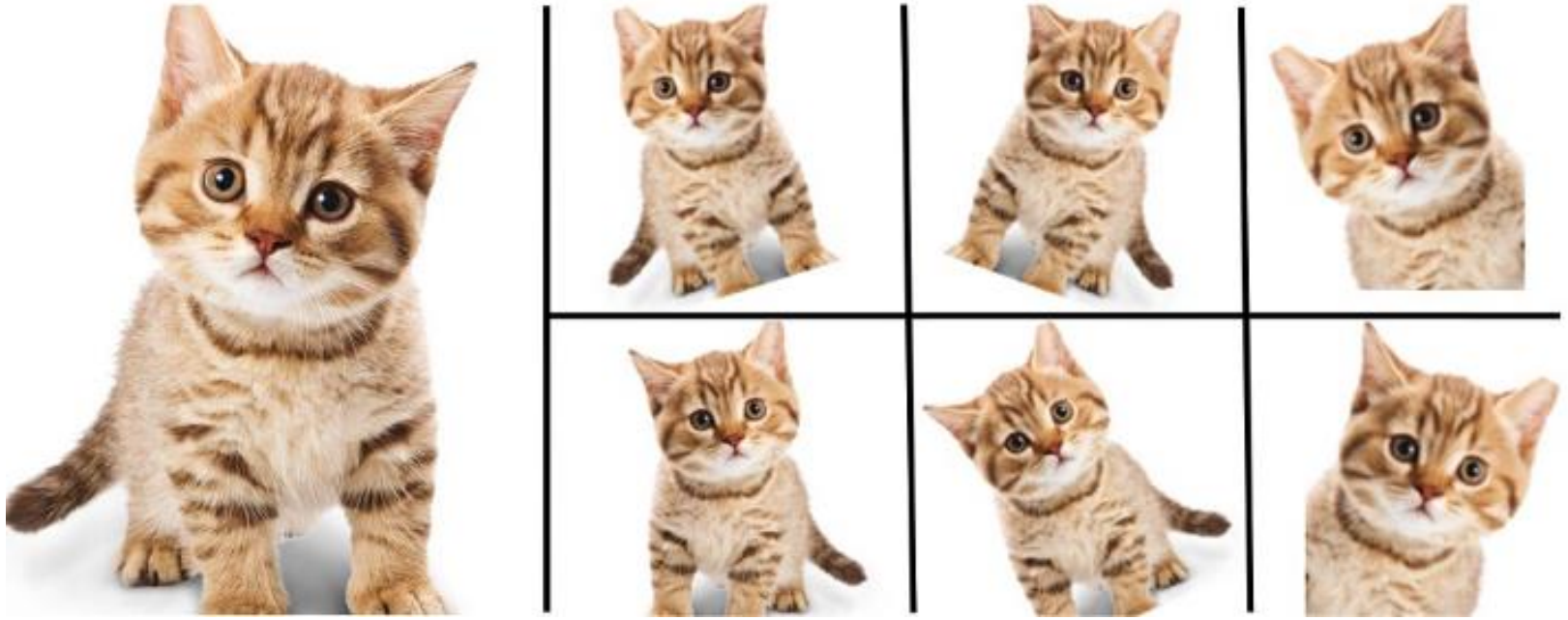


Image from https://wikidocs.net/166068

Intro: Difficulty of Text Data Augmentation

Augmentation

Flipping

Rotating

Cutout

Preserving Semantics

This is a cat

Augmentation

Is this a cat

Semantics Changed

Intro: Previous Text Augmentation – EDA

Make random word-level changes

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

EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks
Wei et al., EMNLP 2019

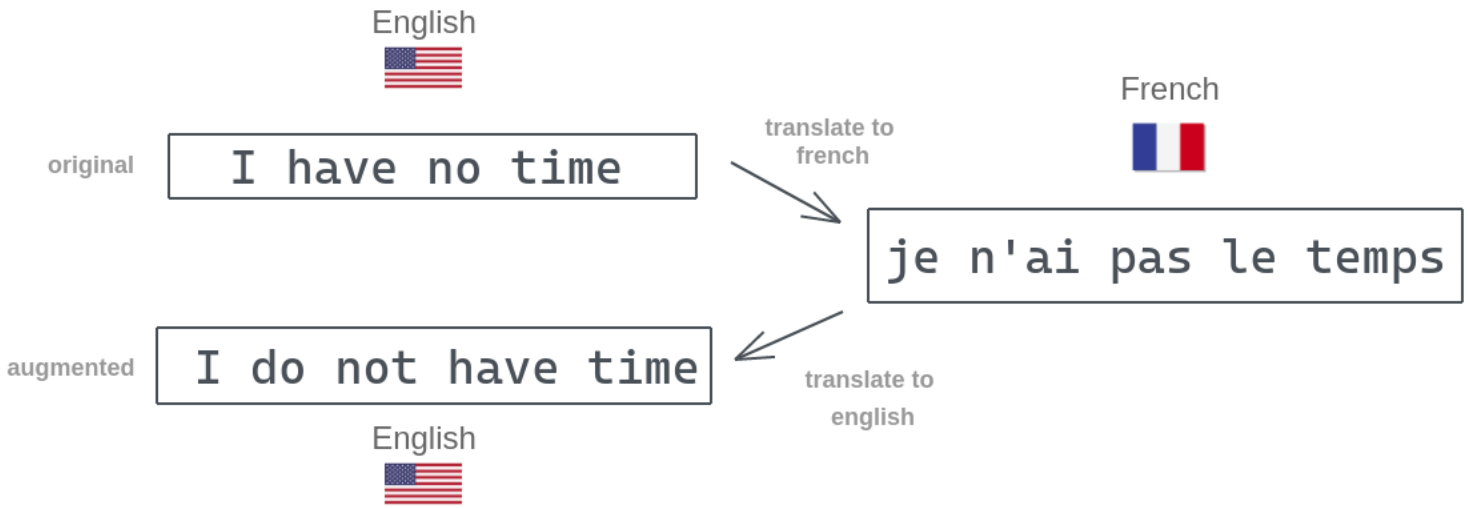
Intro: Previous Text Augmentation – AEDA

Random insertion of punctuation marks

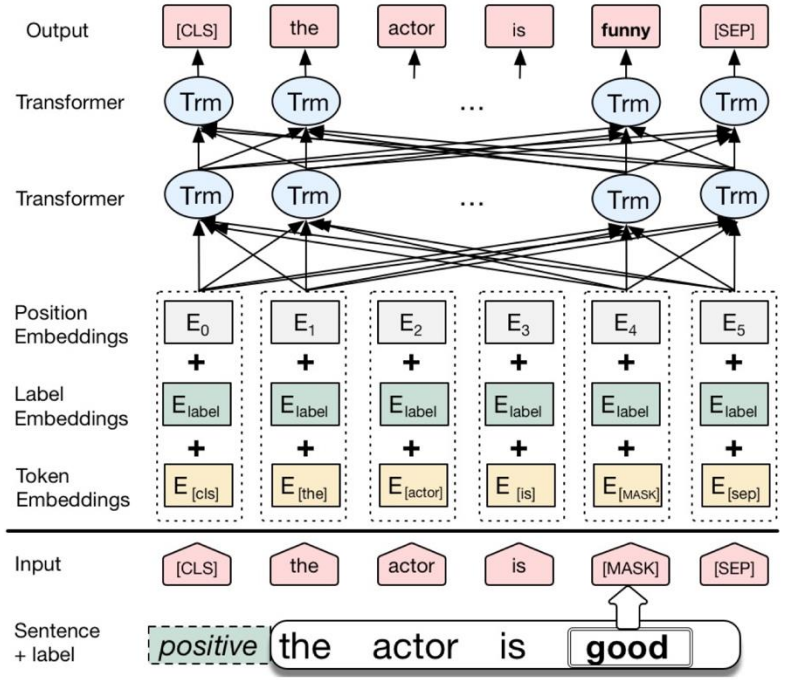
Original	a sad , superior human comedy played out on the back roads of life .
Aug 1	a sad , superior human comedy played out on the back roads ; of life ; .
Aug 2	a , sad . , superior human ; comedy . played . out on the back roads of life .
Aug 3	: a sad ; , superior ! human : comedy , played out ? on the back roads of life .

AEDA: An Easier Data Augmentation Technique for Text Classification
Karimi et al., Findings of EMNLP 2021

Intro: Previous Text Augmentation – Others



Back-Translation
Sennrich et al.



Conditional BERT
Wu et al.

Image from https://amitness.com/2020/02/back-translation-in-google-sheets/

Our Intuition

How can we **improve existing rule-based methods**
while **maintaining its simplicity**?

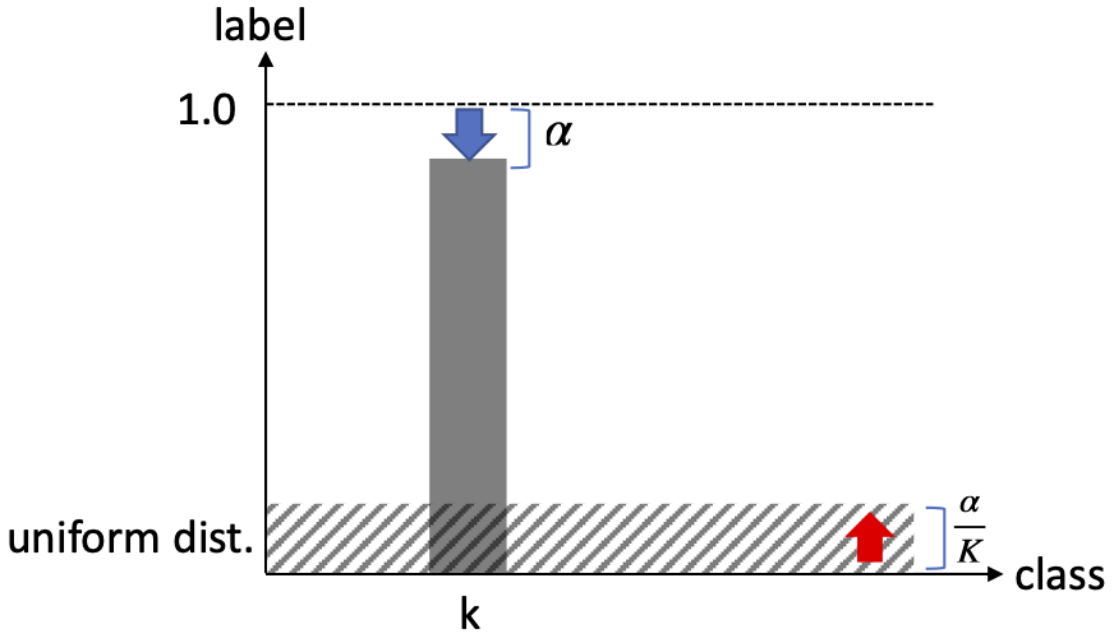
Our Intuition

Rule-based text augmentation methods
make **perturbed data**

Is it okay to assign ground-truth label to such data?
Augmented data should have **less certain label**

Method: Apply Label Smoothing

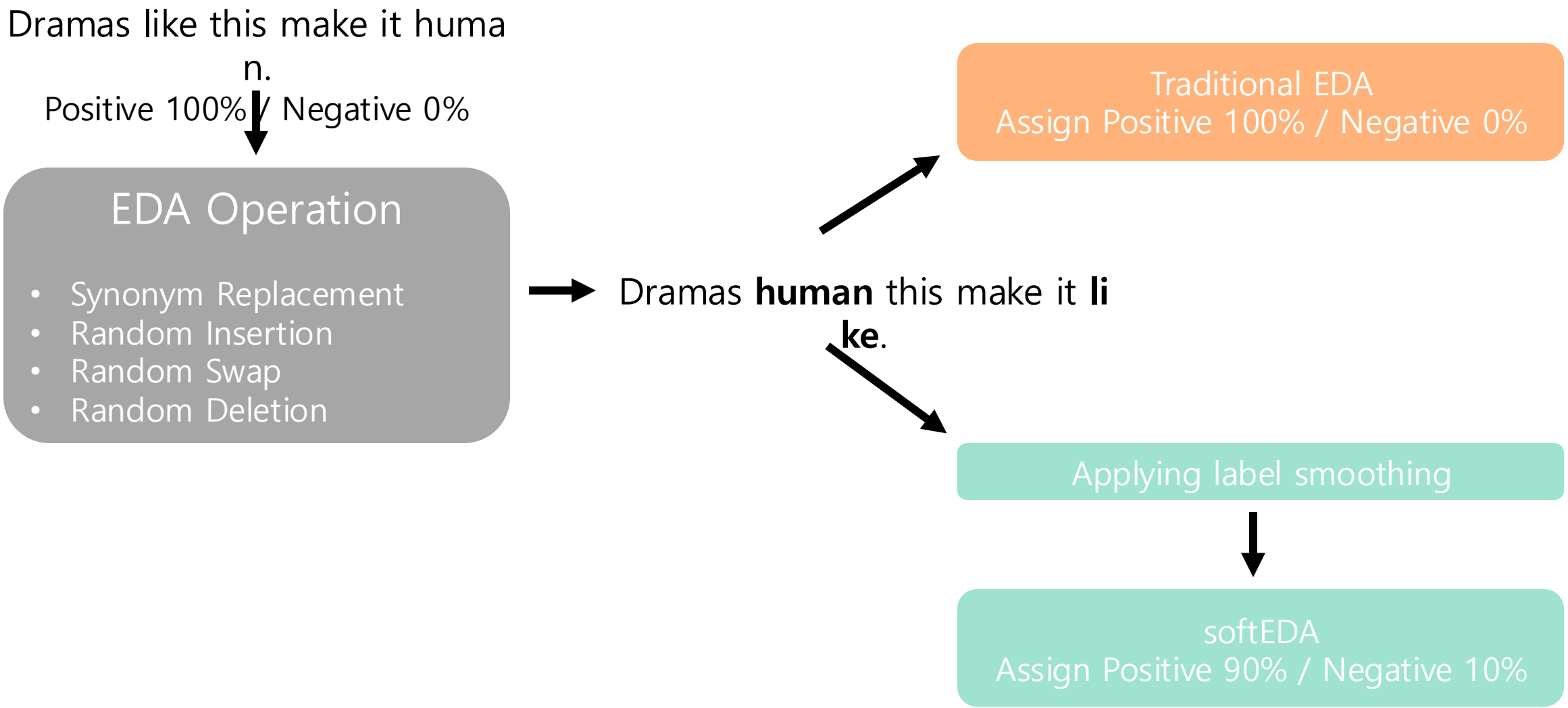
By applying **label smoothing to augmented data**, we can introduce such uncertainty



Rethinking the Inception Architecture for Computer Vision
Szegedy et al., CVPR 2016

Image from https://blog.si-analytics.ai/21

Method: Overall flow



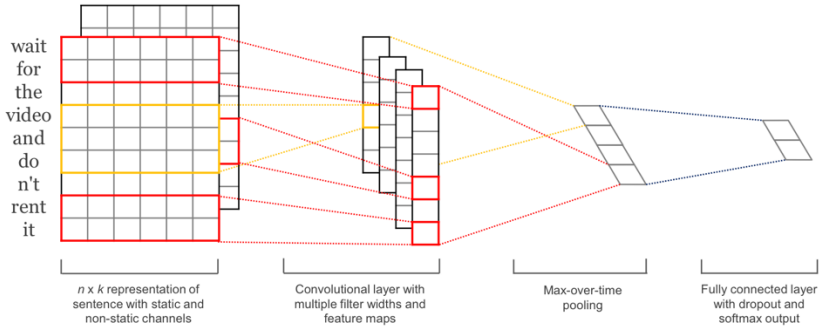
Experiment: Datasets

Evaluated on Seven different text classification dataset

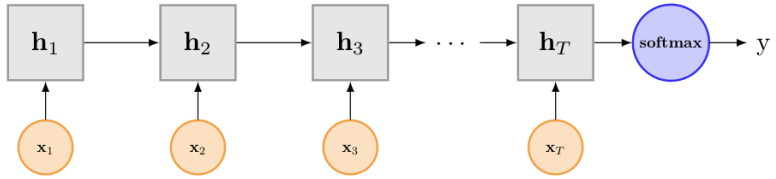
Dataset	Dataset			
	Task	N_{Class}	N_{Train}	N_{Test}
SST2 (Socher et al., 2013)	Sentiment	2	6,919	1,820
CR (Hu & Liu, 2004) (Liu et al., 2015)	Sentiment	2	3,011	752
MR (Pang et al., 2002)	Sentiment	2	9,593	1,067
TREC (Li & Roth, 2002)	Question Type	6	5,452	500
SUBJ (Pang & Lee, 2004)	Subjectivity	2	8,000	2,000
PC (Ganapathibhotla & Liu, 2008)	Pro-Con	2	39,418	4,506
CoLA (Warstadt et al., 2019)	Linguistic Acceptability	2	8,551	527

Experiment: Backbone Models

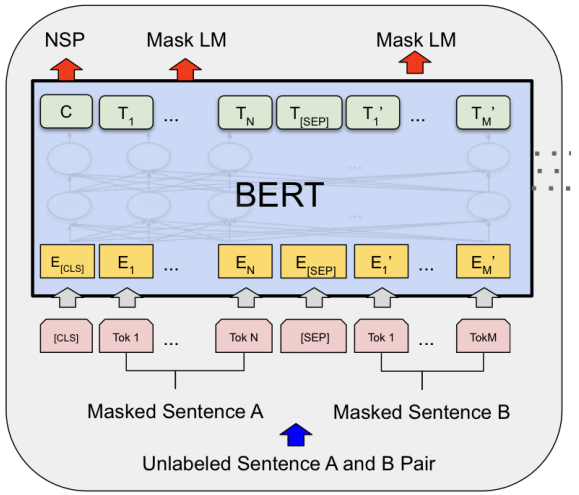
Three different text classification model



CNN-based Model



LSTM-based Model



BERT-based Model

Experiment: Baseline Methods

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

EDA Augmentation

Original	a sad , superior human comedy played out on the back roads of life .
Aug 1	a sad , superior human comedy played out on the back roads ; of life ; .
Aug 2	a , sad . , superior human ; comedy . played . out on the back roads of life .
Aug 3	: a sad ; , superior ! human : comedy , played out ? on the back roads of life .

AEDA Augmentation

Experiment: Results

Model	Dataset						
	SST2	CR	MR	TREC	SUBJ	PC	CoLA
CNN w/o Aug	77.84	77.29	74.94	86.76	90.18	91.62	69.13
w/ EDA	+0.12	-0.02	+0.70	+0.15	+0.05	+0.81	-1.50
w/ AEDA	+0.72	-0.36	-0.40	+0.90	-0.50	+0.04	-2.23
w/ softEDA	**+0.83**	**+0.91**	**+1.84**	**+2.03**	**+0.99**	**+1.29**	**+0.21**
LSTM w/o Aug	75.80	74.26	74.05	86.37	88.84	92.74	69.18
w/ EDA	+0.82	+1.70	+0.70	-3.24	+1.69	+0.49	-0.07
w/ AEDA	**+2.97**	+0.28	+0.49	-1.37	+0.64	-0.15	-0.37
w/ softEDA	+2.59	**+2.90**	**+1.41**	**+1.95**	**+2.18**	**+0.60**	**+0.18**
BERT w/o Aug	89.74	89.08	84.28	95.47	96.18	93.44	75.38
w/ EDA	+0.71	-0.41	-0.92	+0.51	-0.35	+0.58	-0.45
w/ AEDA	+0.22	+1.84	**+0.19**	-0.67	-0.30	-0.15	-0.34
w/ softEDA	**+0.83**	**+2.10**	**+0.19**	**+1.17**	**+0.15**	**+0.67**	**+1.50**

Conclusion

Improving simple augmentation methods while maintaining its simplicity!

Thank you for listening!

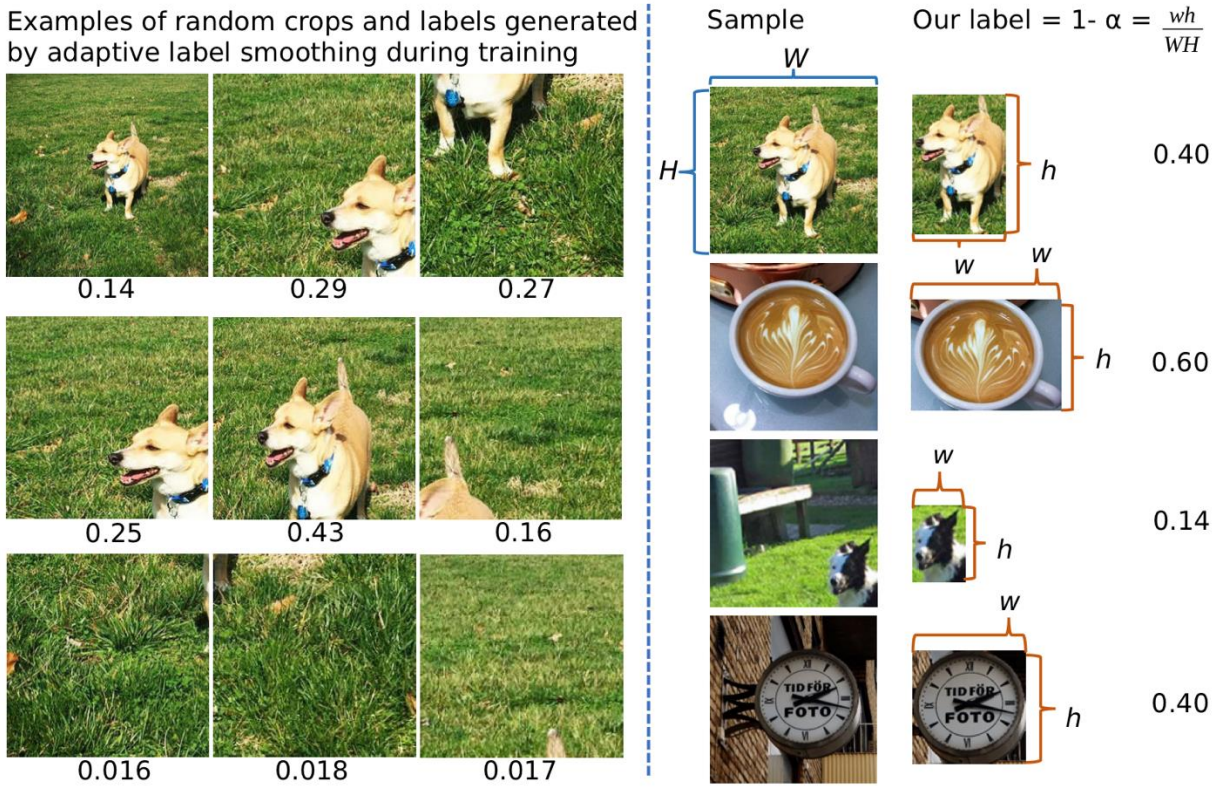
Appendix: Smoothing Factor

Heuristic search of smoothing factor

Model	Dataset						
	SST2	CR	MR	TREC	SUBJ	PC	CoLA
CNN w/o Augmentation	77.84	77.29	74.94	86.76	90.18	91.62	69.13
w/ EDA	77.96	77.27	75.64	86.91	90.23	92.43	67.63
w/ AEDA	78.56	76.93	74.54	87.66	89.68	91.66	66.90
w/ softEDA $\alpha = 0.1$	**78.67**	76.13	74.88	87.50	90.33	92.61	68.05
w/ softEDA $\alpha = 0.15$	77.07	75.48	73.99	**88.79**	89.88	**92.91**	68.79
w/ softEDA $\alpha = 0.2$	77.90	**78.20**	74.75	87.73	90.13	92.35	68.79
w/ softEDA $\alpha = 0.25$	78.10	77.91	75.18	88.01	90.67	92.49	68.97
w/ softEDA $\alpha = 0.3$	75.48	76.10	**76.78**	87.54	**91.17**	92.69	**69.34**
LSTM w/o Augmentation	75.80	74.26	74.05	86.37	88.84	92.74	69.18
w/ EDA	76.62	75.96	74.75	83.13	90.53	93.23	69.11
w/ AEDA	**78.77**	74.54	74.54	85.00	89.48	92.59	68.81
w/ softEDA $\alpha = 0.1$	78.39	**77.16**	73.62	87.46	88.94	93.12	69.18
w/ softEDA $\alpha = 0.15$	74.15	72.99	73.71	86.68	89.93	**93.34**	69.18
w/ softEDA $\alpha = 0.2$	76.09	75.73	73.01	**88.32**	89.53	93.07	69.00
w/ softEDA $\alpha = 0.25$	77.23	75.48	71.48	86.84	90.48	92.92	67.34
w/ softEDA $\alpha = 0.3$	77.18	75.61	**75.46**	86.09	**91.02**	92.63	**69.36**
BERT w/o Augmentation	89.74	89.08	84.28	95.47	96.18	93.44	75.38
w/ EDA	90.45	88.67	83.36	95.98	95.83	94.02	74.93
w/ AEDA	89.96	90.92	**84.47**	94.80	95.88	93.29	75.04
w/ softEDA $\alpha = 0.1$	89.63	89.37	83.18	94.02	**96.33**	93.87	76.72
w/ softEDA $\alpha = 0.15$	89.52	89.74	83.82	95.00	95.68	93.43	75.40
w/ softEDA $\alpha = 0.2$	89.62	**91.18**	**84.47**	95.20	96.23	93.87	76.19
w/ softEDA $\alpha = 0.25$	89.51	**91.18**	83.36	**96.64**	96.08	**94.11**	**76.88**
w/ softEDA $\alpha = 0.3$	**90.57**	88.18	82.48	94.69	96.18	**94.11**	75.61

Appendix: Similar Work in Vision Field

Applying adaptive smoothing factor after cropping



One Size Doesn't Fit All: Adaptive Label Smoothing
Krothapalli and Abbott, arXiv 2020