

Tasks

Hee Hwang and Sudarshan Raghavan

April 30, 2021

1 Baseline

ACL : Citation Intent Classification

Hyper: HyperPartisan News Detection

RCT : Randomized Controlled Trials

1.1 Baseline

Task	train	dev	test	Classes
ACL	1688	114	139	6
Hyper	516	64	65	2
RCT-sample	500	30212	30135	5
RCT-200k	180040	30212	30135	5

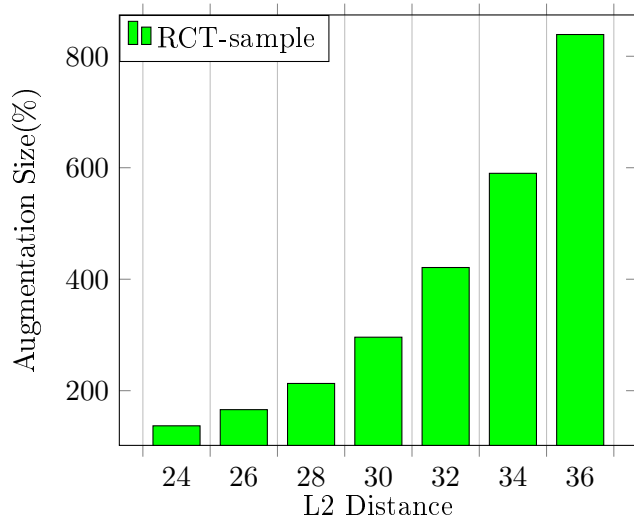
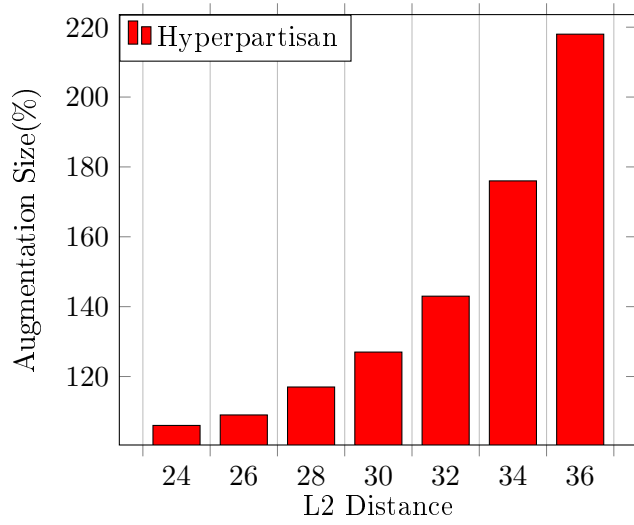
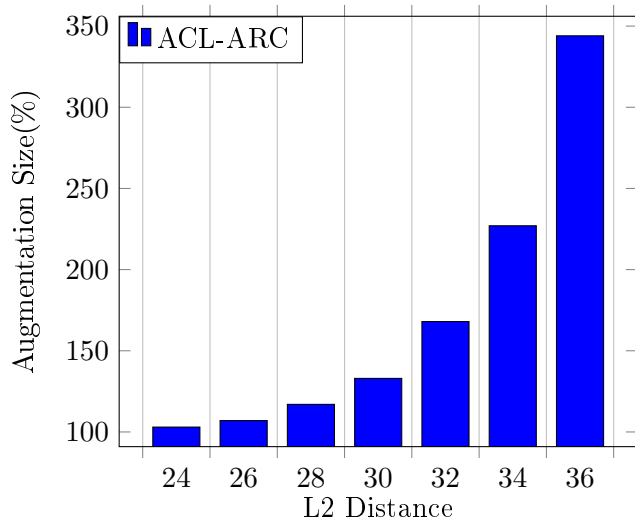
2 Tables & Plots

2.1 Augmentation by Distance

2.1.1 Size Table

Max Distance	ACL	Hyper	RCT _{sample}
Baseline	1688 (100%)	516 (100%)	500(100%)
24	1746 (103%)	551 (106%)	686(137%)
26	1815 (107%)	567 (109%)	831(166%)
28	1981 (117%)	606 (117%)	1065(213%)
30	2253 (133%)	656 (127%)	1484(296%)
32	2842 (168%)	742 (143%)	2105(421%)
34	3848 (227%)	911 (176%)	2952(590%)
36	5819 (344%)	1127 (218%)	4196(839%)

2.1.2 Size Plot

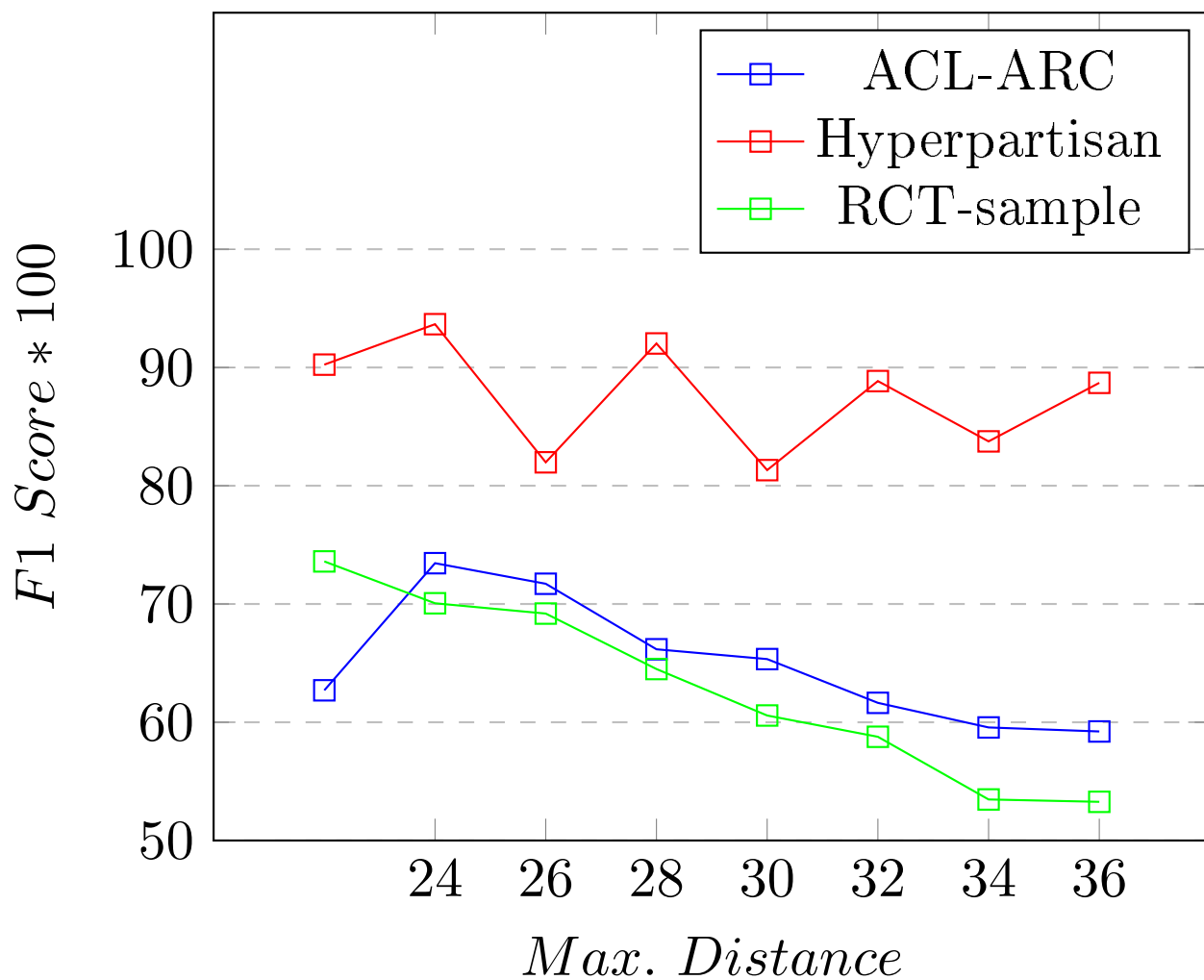


2.1.3 F1 Table

Max Distance	ACL	Hyper	RCT _{sample}
Baseline	62.70	90.24	73.60
24	73.45	93.66	70.06
26	71.71	81.98	69.19
28	66.17	92.03	64.50
30	65.34	81.32	60.57
32	61.64	88.85	58.77
34	59.56	83.75	53.47
36	59.22	88.70	53.27

2.1.4 F1 Plot

Augmentation by Maximum Distance



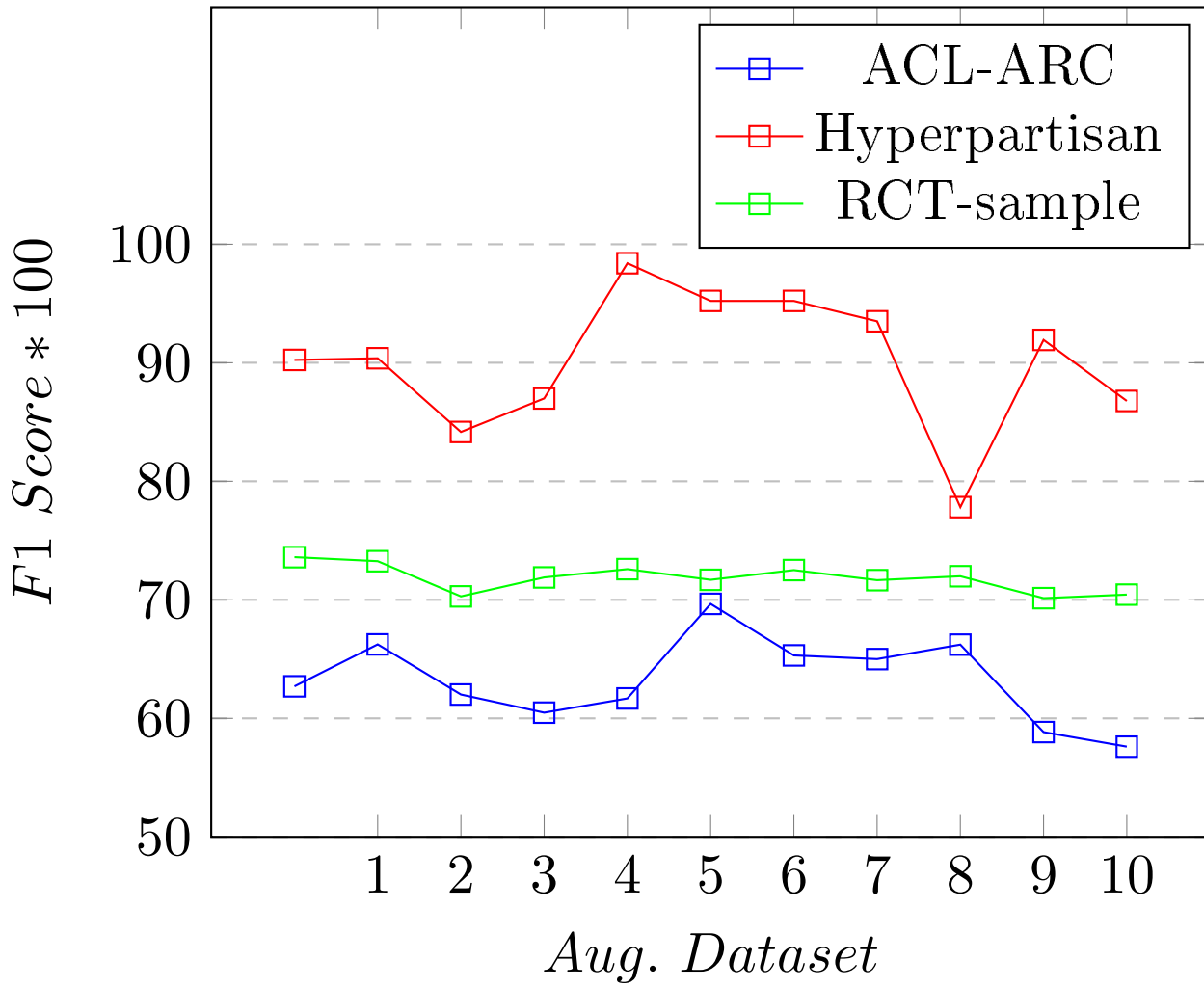
2.2 Augmentation by Fixed Size

2.2.1 F1 Table

Dataset #	ACL	Hyper	RCT _{sample}	Interval
Baseline	62.70	90.24	73.60	-
1	66.24	90.38	73.26	0-3%
2	62.01	84.16	70.29	3-6%
3	60.48	86.99	71.89	6-9%
4	61.68	98.40	72.59	9-12%
5	69.66	95.22	71.69	12-15%
6	65.31	95.22	72.51	15-18%
7	65.00	93.50	71.66	18-21%
8	66.22	77.82	71.99	21-24%
9	58.84	91.93	70.14	24-27%
10	57.61	86.78	70.44	27-30%

2.2.2 F1 Plot

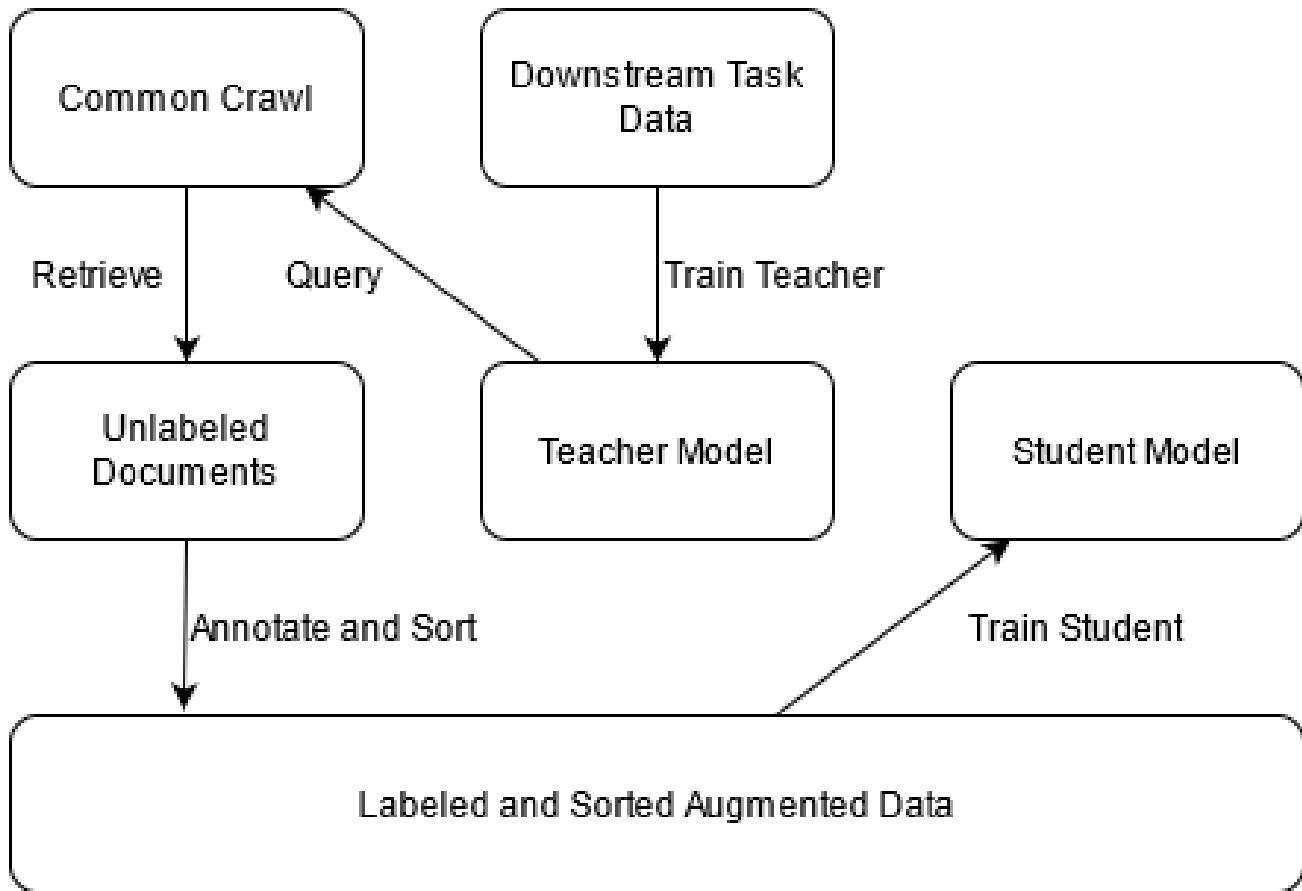
Augmentation by Size (Non-cumulative)



3 Baseline models:

- An off-the-shelf RoBERTa model that has been finetuned to perform classification for each of the downstream tasks

4 Augmentation Model



5 Algorithm

1. Extract failed test examples from the baseline model
2. Retrieve passages/sentences from Common Crawl
3. Apply augmentation strategy (i)-(iii)
4. Augment all the labelled CC data to the training data
5. Retrain RoBERTa on the augmented training set

6 Augmentation Strategies

- Strategy (i)
Use baseline model (Teacher) to perform unsupervised labelling on retrieved CC data
- Strategy (ii)
Using a task specific binary classifier, filter out retrieved CC data that is "out-domain"
Use baseline model (Teacher) to perform unsupervised labelling on the filtered "in-domain" CC data
- Strategy (iii)
Using a task specific binary classifier, filter out retrieved CC data that is "out-domain"
Use ground truth labels of failed test examples and assign labels to the filtered "in-domain" CC data