Tasks

Hee Hwang and Sudarshan Raghavan

May 12, 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

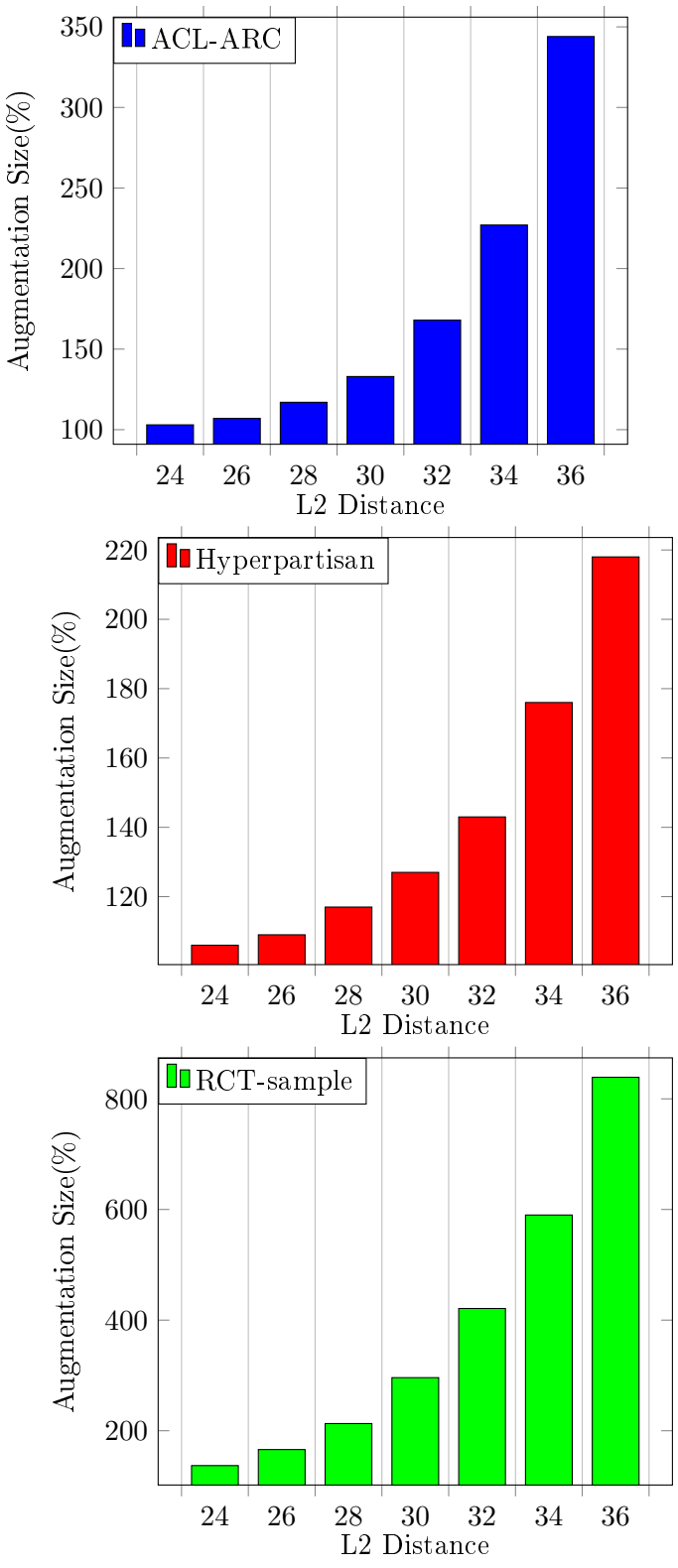
2 Tables & Plots

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%)

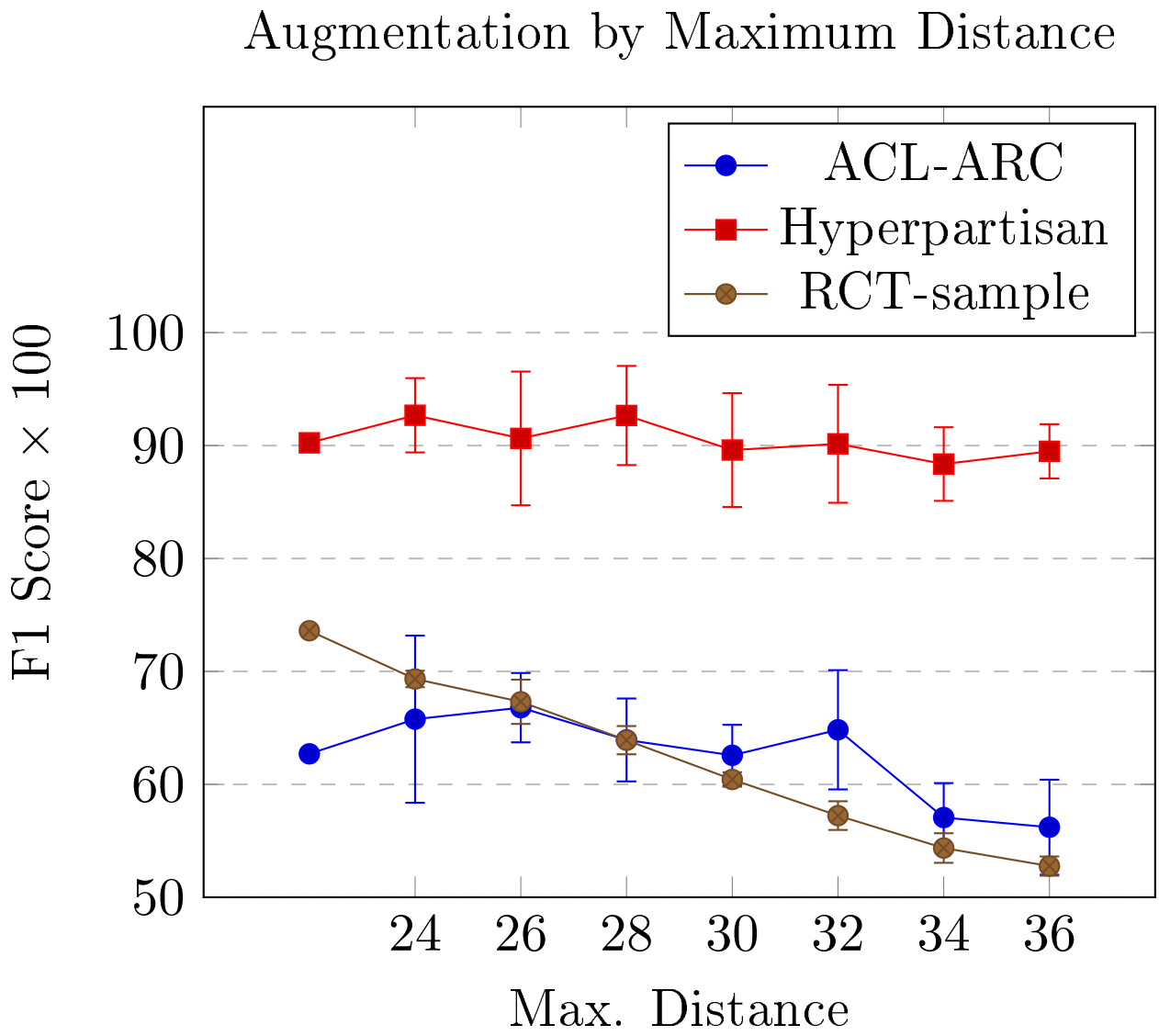
2.1.2 Size Plot



2.1.3 F1 Table

Max Distance	ACL	Hyper	RCT_{sample}
Baseline	62.70	90.24	73.60
24	65.77(± 7.40)	92.68(± 3.29)	69.33(± 0.73)
26	66.79(± 3.07)	90.63(± 5.92)	67.31(± 1.96)
28	63.93(± 3.67)	92.66(± 4.39)	63.91(± 1.25)
30	62.57(± 2.70)	89.60(± 5.04)	60.43(± 0.64)
32	64.83(± 5.28)	90.15(± 5.22)	57.23(± 1.27)
34	57.06(± 3.05)	88.36(± 3.26)	54.36(± 1.31)
36	56.20(± 4.21)	89.49(± 2.40)	52.76(± 0.85)

2.1.4 F1 Plot

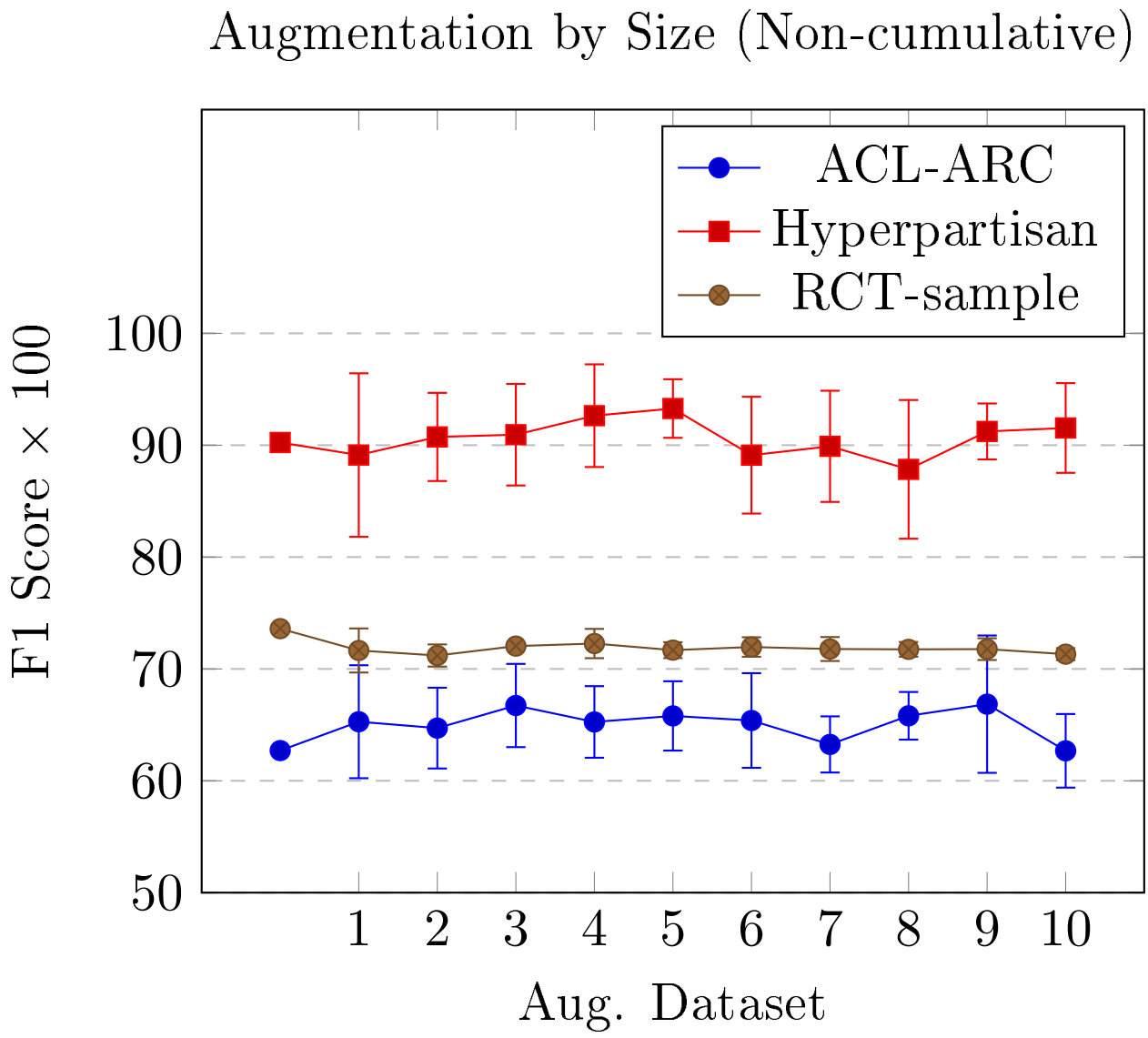


2.2 Augmentation by Fixed Size

2.2.1 F1 Table

Aug Data	ACL	Hyper	RCT_{sample}
Baseline	62.70	90.24	73.60
1	65.28(± 5.05)	89.12(± 7.31)	71.65(± 1.97)
2	64.71(± 3.61)	90.74(± 3.94)	71.20(± 0.99)
3	66.73(± 3.72)	90.94(± 4.54)	72.04(± 0.20)
4	65.26(± 3.20)	92.64(± 4.59)	72.27(± 1.31)
5	65.80(± 3.10)	93.28(± 2.62)	71.68(± 0.71)
6	65.39(± 4.23)	89.11(± 5.22)	71.96(± 0.87)
7	63.25(± 2.51)	89.90(± 4.97)	71.78(± 1.07)
8	65.81(± 2.13)	87.84(± 6.20)	71.75(± 0.66)
9	66.85(± 6.14)	91.23(± 2.50)	71.77(± 0.97)
10	62.68(± 3.29)	91.54(± 4.01)	71.32(± 0.54)

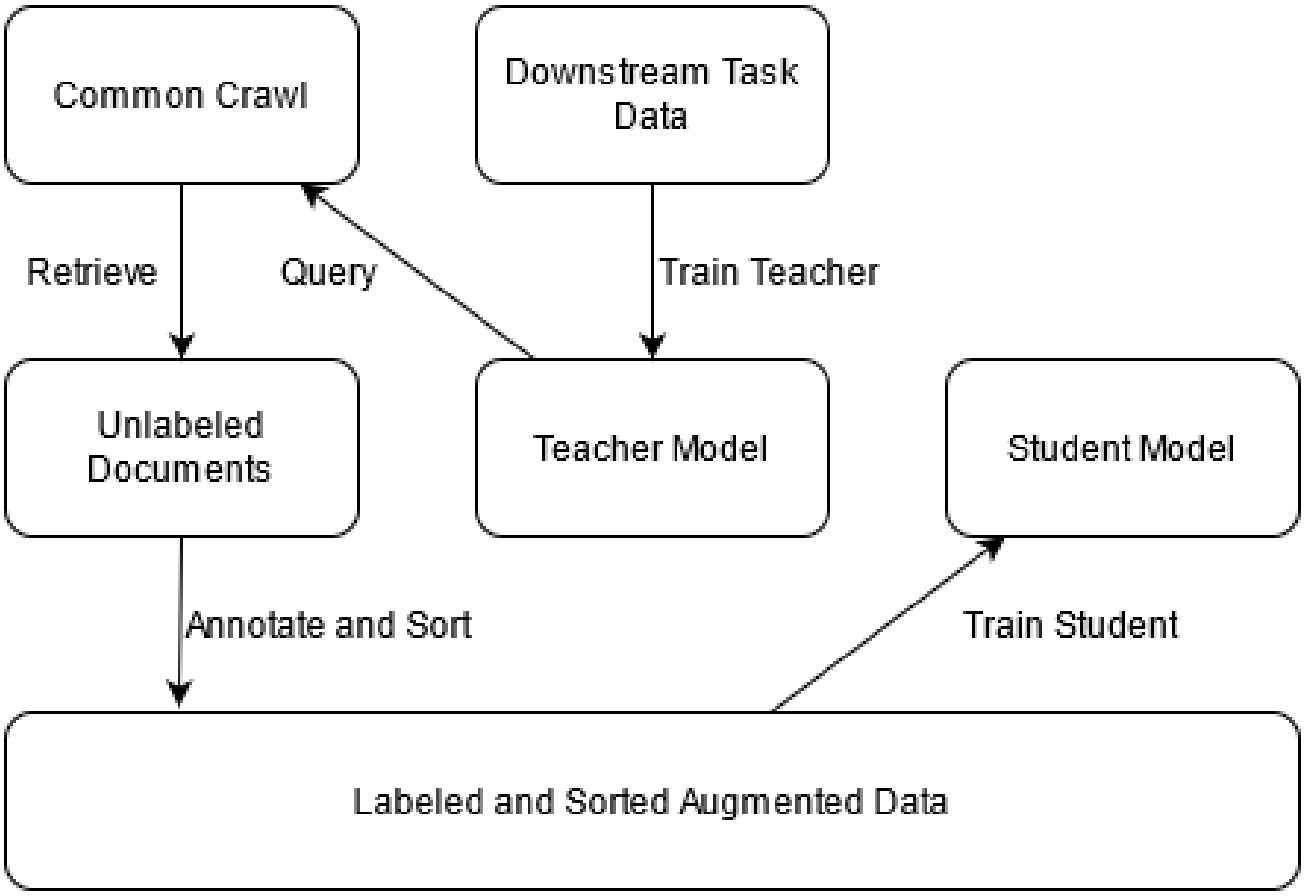
2.2.2 F1 Plot



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