A Pilot Study of Domain Adaptation Effect for Neural Abstractive Summarization

Northeastern University
College of Computer and Information Science

Xinyu hua and Lu Wang
hua.x@husky.neu.edu luwang@ccs.neu.edu

Introduction

Neural abstractive summarization proves to be effect in news abstracts generation. But such seq2seq requires large amounts of training data. For certain domains this is difficult. We investigate the following two research questions as an initial study on how to solve this issue:

- *Can we leverage out-of-domain abstracts and extractive summaries to help train a neural summarization model for a new domain?*
- *What information can be transferred and what are the limitations?*

Two aspects to study:

- *Effect of parameter initialization with extractive summaries.*
- *Feasibility of domain adaptation and what specific information can be transferred to a new domain.*

Data and Experiment Setup

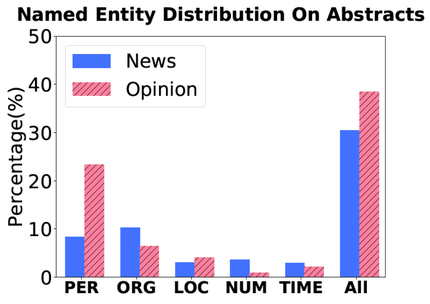
Primary Data Source: The New York Times Annotated Corpus (Sandhaus, 2008). The type of document is determined by its taxonomy tags coming with the dataset. The resultant dataset contains 100,824 news stories and 51,214 opinion articles.

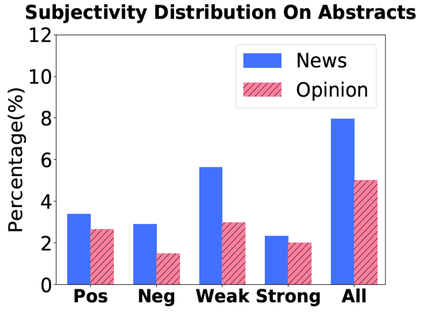
Domains:

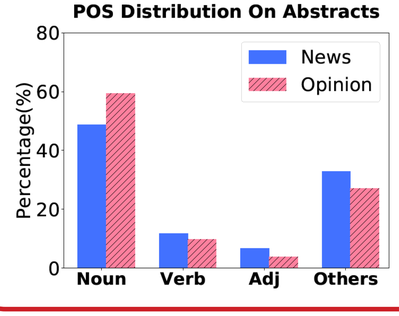
- ***News***: *News,*
- ***Opinion***: *Editorial, Opinion, Features*

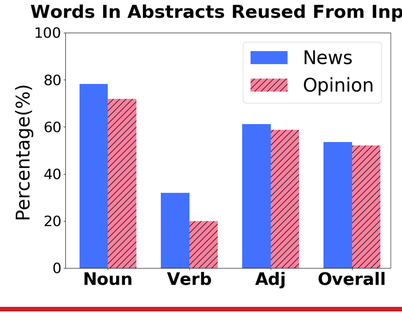
Pre-training Dataset: One million lead paragraph and article pairs.

Training Setup: train (75%), valid (15%), and test(10%). Experiments are conducted in the following setups.





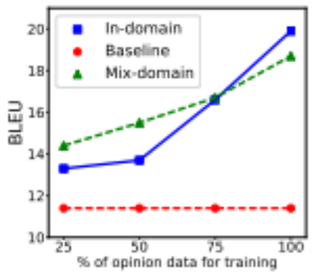


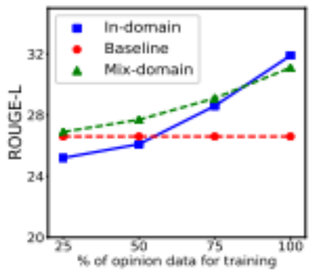


Results

Effect of pre-training: Baseline 1 outputs the first sentence in original document. Baseline 2 selects first 22 tokens (news) and 15 tokens (opinion). The pre-training step improves performance for NEWS, whereas the performance on OPINION remains roughly the same.

	R-2	R-L	BLEU	Avg. Len.
Test on News				
BASELINE1	23.5	35.4	19.9	28.94
BASELINE2	19.5	30.1	19.5	22.00
IN-DOMAIN	23.3	34.1	21.3	22.08
IN-DOMAIN + pretrain	**24.2**	**34.5**	**22.4**	21.59
Test on Opinion				
BASELINE1	17.9	26.6	11.4	28.18
BASELINE2	12.9	20.5	11.7	15.00
IN-DOMAIN	19.8	**31.9**	**19.9**	14.60
IN-DOMAIN + pretrain	**19.9**	31.8	19.4	14.22





	Train	Test
IN-DOMAIN	News/Opinion	News/Opinion
OUT-OF-DOMAIN	News	Opinion
MIX-DOMAIN	News + Opinion	Opinion

Effect of Domain Adaptation: when In-domain data is insufficient, Mix-domain setup yields better performance. Sample summaries are also displayed:

	Abstract
Human	stephen holden reviews carnegie hall concert celebrating music of judy garland. singers include her daughter, lorna luft.
Out-of-Domain	article discusses possibility of carnegie hall in carnegie hall golf tournament.
Mix-Domain	stephen holden reviews performance by jazz singer celebration by rainbow and garland at carnegie, part of tribute hall.
Human	janet maslin reviews john grisham book the king of torts .
Out-of-Domain	interview with john grisham of legal thriller is itself proof for john grisham 376 pages.
Mix-Domain	janet maslin reviews book the king of torts by john grisham .
Human	anthony tommasini reviews 23d annual benefit concert of richard tucker music foundation , featuring members of metropolitan opera orchestra led by leonard slatkin .
Out-of-Domain	final choral society and richard tucker music foundation , on sunday night in [UNK] fisher hall , will even longer than substantive 22d gala last year .
Mix-Domain	anthony tommasini reviews 23d annual benefit concert of benefit of richard tucker music.

Analysis

- *The model trained for news is better at generating tokens in the input. The model trained for opinion is better at generating new words not in the input.*
- *Attention distribution on word types is consistent with that in summary, thus the identification of summary-worthy named entities might be transferable.*
- *Ability to select salient content is largely kept for Mix-domain training.*
- *Out-of-domain style could not be learned in this setup.*

Src -> Trt	IN-DOMAIN *News -> News*	OUT-OF-DOMAIN *News -> Opin*	MIX-DOMAIN *News + Opin -> Opin*
PER	7.9%	8.7%	15.1%
ORG	10.9%	6.9%	8.2%
All NEs	26.7%	23.6%	31.6%
Noun	41.2%	36.2%	43.3%
Verb	10.3%	6.7%	5.5%
Positive	5.6%	5.1%	4.5%
Negative	2.5%	2.2%	2.1%

Conclusion and Future Direction

- We investigated domain adaptation for abstractive neural summarization. Experimental results showed that pre-training model with extractive summaries helps.
- By analyzing the attention weight distribution over input tokens, we found the model was capable to select salient information even trained on out-of-domain data.
- Our next step is to develop domain adaptation techniques to allow a summarization system to learn content selection from out-of-domain data while acquiring language generating behavior with in-domain data.