Efficient Few-Shot Fine-Tuning for Opinion Summarization

Presented by:
Kushan Hewapathirana – 229333P

Authors:

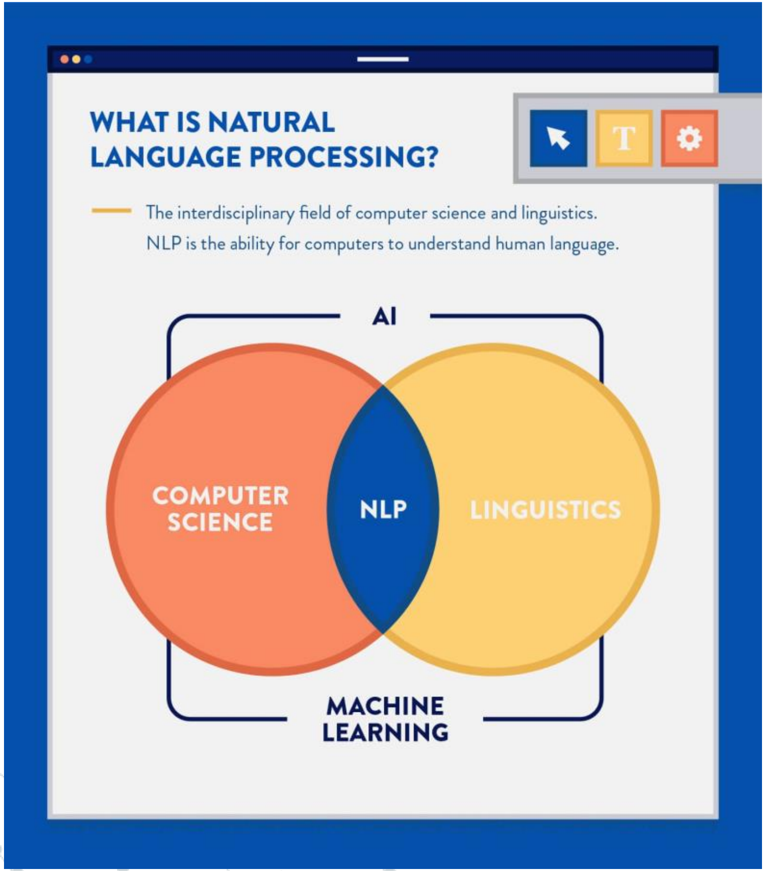
1. **Arthur Bražinskas (h index: 5)**
2. **Ramesh Nallapati (h index: 28)**
3. **Mohit Bansal (h index: 59)**
4. **Markus Dreyer (h index: 14)**

Published in: NAACL : July, 2022

APPLICATION DOMAIN:

Natural Language Processing – Opinion Summarization

Introduction to Natural Language Processing



The interdisciplinary field of computer science and linguistics. NLP is the ability for computers to understand human language.

- Speech recognition
- Part of speech tagging
- Word sense disambiguation
- Named entity recognition
- Co-reference resolution
- Sentiment analysis
- Natural language generation

Introduction to Opinion Summarization

Generating a concise summary of a set of reviews or opinions on a particular product or service

Goal is to capture the overall sentiment and key aspects of the reviews in a condensed form

Useful for decision-making, such as when deciding whether to purchase a product or use a service.

Techniques and Methods used for Opinion Summarization

•**Supervised learning**: Training a model on a labeled dataset of opinionated text and corresponding summaries, and then using the model to generate summaries for new text.

•**Unsupervised learning**: Using clustering, topic modeling, and other unsupervised techniques to identify the main topics and sentiments expressed in the text, and then generating a summary based on those topics and sentiments.

•**Deep learning**: Using neural networks, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to learn the patterns and relationships in the text and generate summaries based on those patterns and relationships

•**Language models**: Using pre-trained language models, such as BERT, to generate summaries based on the input text

•**Evaluation metrics**: There are several evaluation metrics used to assess the quality of opinion summarization, including ROUGE, BLEU

PROBLEM STATEMENT:

Lack of large annotated datasets for opinion summarization

Challenges in Opinion Summarization

Large annotated datasets of reviews paired with reference summaries are not available.

Such datasets would be expensive to create and requires fine-tuning methods robust to overfitting on small datasets.

Pre-trained models are often not accustomed to the specifics of customer reviews and, after fine-tuning, yield summaries with disfluencies and semantic mistakes.

Bridging the Research Gap: Unique Contribution

Large annotated datasets of reviews paired with reference summaries are not available and would be expensive to create

- Efficient few-shot method based on adapters and self-supervised pre-training, which can easily store in-domain knowledge and is robust to overfitting on small datasets.

Generically pre-trained models are not accustomed to the specifics of customer reviews and, after fine-tuning, yield summaries with disfluencies and semantic mistakes

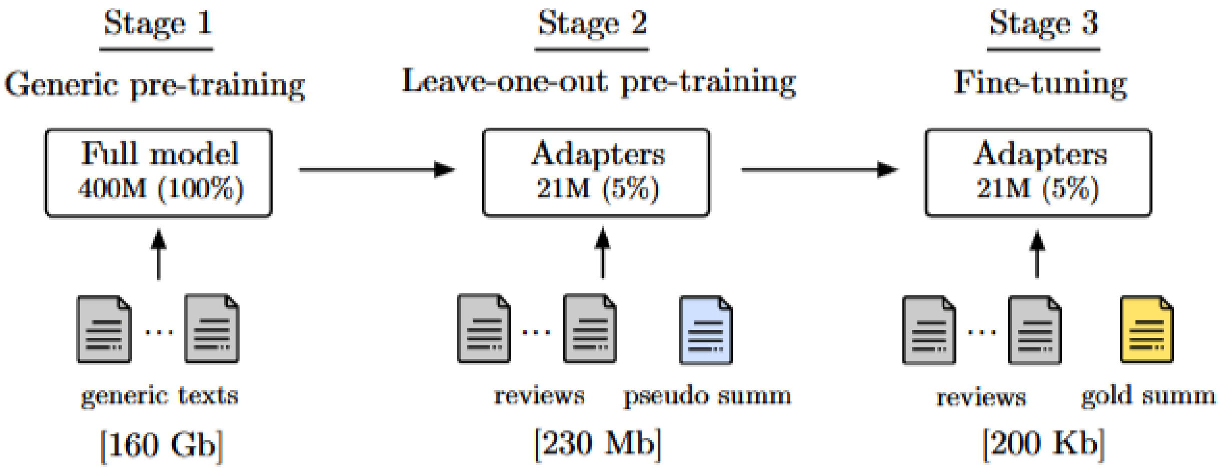
- Adding adapters and pretraining them in a task-specific way on a large corpus of unannotated customer reviews, using held-out reviews as pseudo summaries.
- This self-supervised adapter pre-training improves summary quality over standard fine-tuning.

Summary personalization

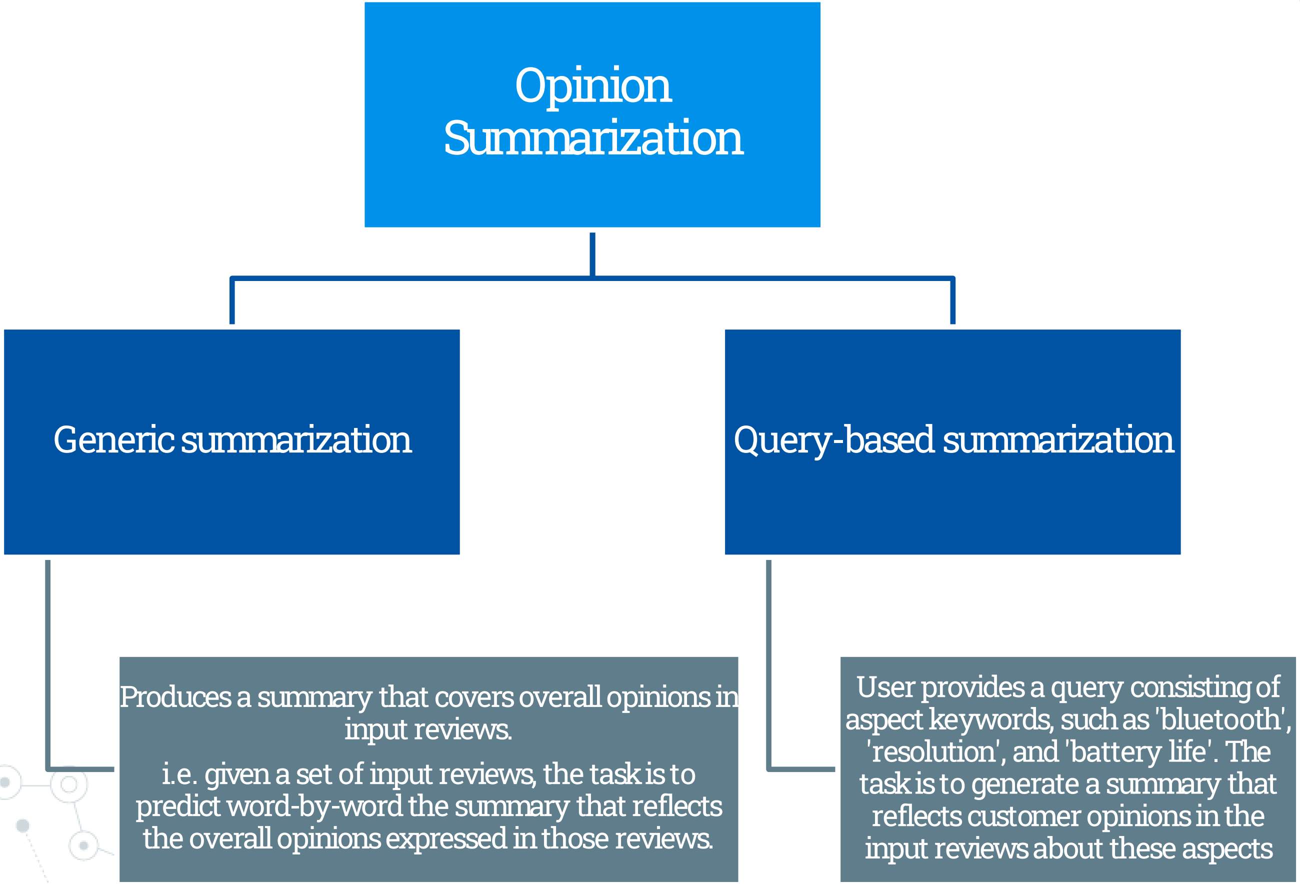
- Conditioning on aspect keyword queries, automatically created from generic datasets.

Methodology:
Efficient Few-Shot Fine-Tuning for Opinion Summarization

Proposed Approach



Opinion Summarization Tasks



Opinion Summarization

Generic summarization

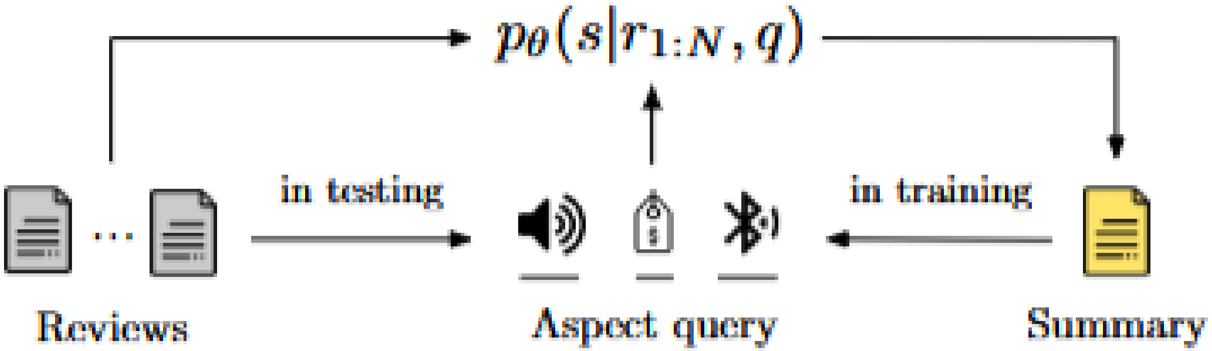
Query-based summarization

Produces a summary that covers overall opinions in input reviews.

i.e. given a set of input reviews, the task is to predict word-by-word the summary that reflects the overall opinions expressed in those reviews.

User provides a query consisting of aspect keywords, such as 'bluetooth', 'resolution', and 'battery life'. The task is to generate a summary that reflects customer opinions in the input reviews about these aspects

Query-based Summarization



Utilized Model: BART

- BART is a denoising autoencoder that is pre-trained on a large corpus of generic texts.

- The encoder and decoder of BART are used for the input and output of the summarization task, respectively.

- The paper fine-tunes BART on a small annotated dataset of reviews and summaries for opinion summarization.

Adapters

- Adapters are tiny neural networks that are optimized during training while the pre-trained model remains frozen.

- The paper proposes to use adapters for opinion summarization to store in-domain knowledge and avoid overfitting on small datasets.

- The adapters are injected into the transformer layers of the pre-trained model.

- The paper fine-tunes the adapters on a small annotated dataset of reviews and summaries for opinion summarization.

Self-supervised pre-training

- The paper proposes a self-supervised pre-training method for adapters on a large corpus of unannotated customer reviews.

- The held-out reviews are used as pseudo summaries for the self-supervised pre-training.

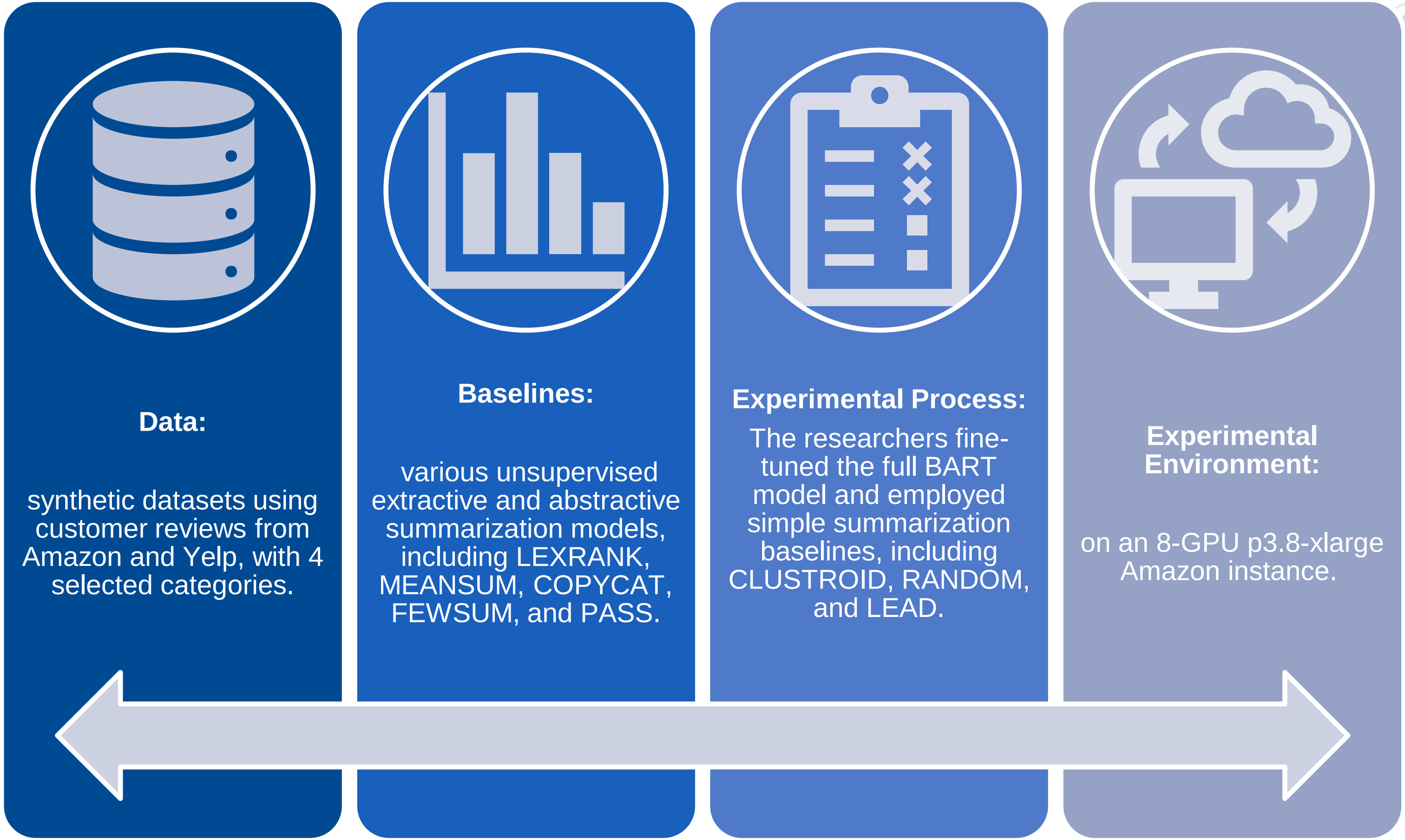
- The adapters are pre-trained in a task-specific way to store in-domain knowledge.

- The paper fine-tunes the adapters on a small annotated dataset of reviews and summaries for opinion summarization.

- The self-supervised adapter pre-training improves summary quality over standard fine-tuning.

EXPERIMENTS AND RESULTS

Experimental Setup



Data:

synthetic datasets using customer reviews from Amazon and Yelp, with 4 selected categories.

Baselines:

various unsupervised extractive and abstractive summarization models, including LEXRANK, MEANSUM, COPYCAT, FEWSUM, and PASS.

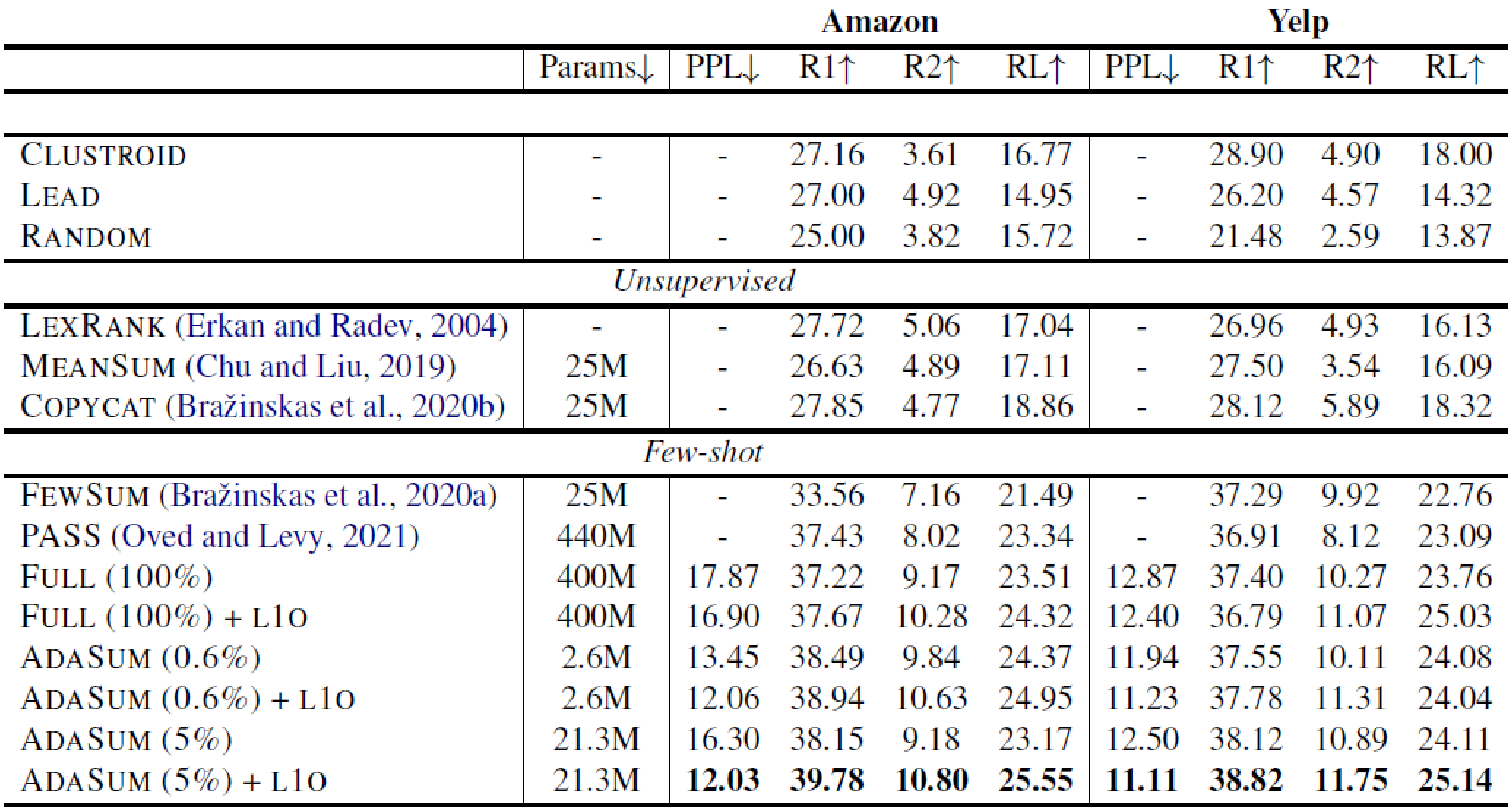
Experimental Process:

The researchers fine-tuned the full BART model and employed simple summarization baselines, including CLUSTROID, RANDOM, and LEAD.

Experimental Environment:

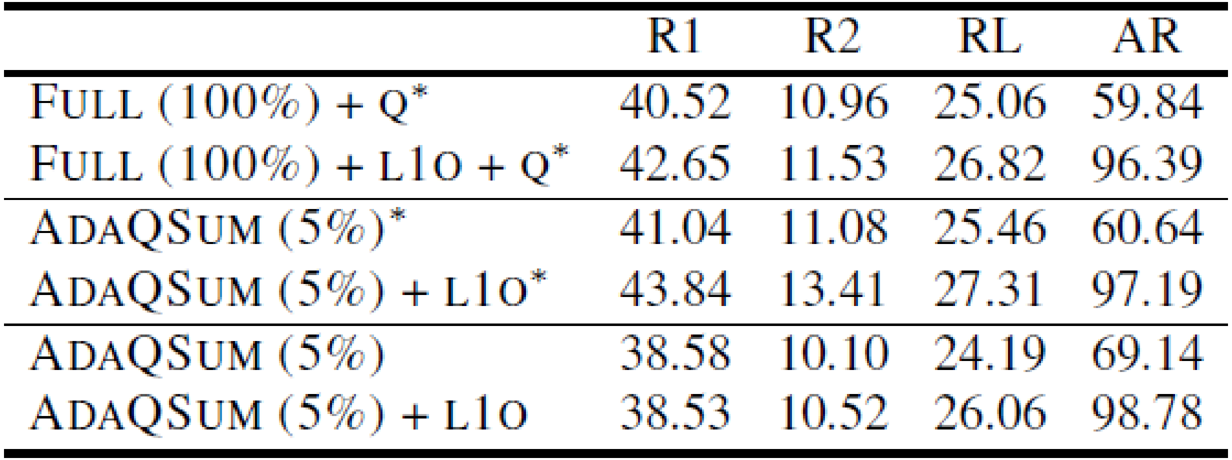
on an 8-GPU p3.8-xlarge Amazon instance.

Test Results - Generic Review Summarization



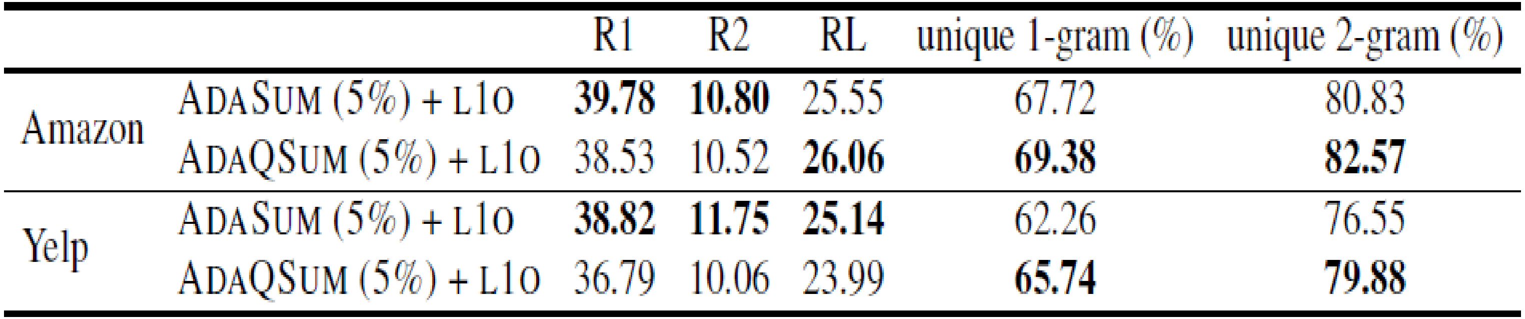
	Amazon					Yelp			
	Params↓	PPL↓	R1↑	R2↑	RL↑	PPL↓	R1↑	R2↑	RL↑
Clustroid	-	-	27.16	3.61	16.77	-	28.90	4.90	18.00
Lead	-	-	27.00	4.92	14.95	-	26.20	4.57	14.32
Random	-	-	25.00	3.82	15.72	-	21.48	2.59	13.87
Unsupervised									
LexRank (Erkan and Radev, 2004)	-	-	27.72	5.06	17.04	-	26.96	4.93	16.13
MeanSum (Chu and Liu, 2019)	25M	-	26.63	4.89	17.11	-	27.50	3.54	16.09
Copycat (Bražinskas et al., 2020b)	25M	-	27.85	4.77	18.86	-	28.12	5.89	18.32
Few-shot									
FewSum (Bražinskas et al., 2020a)	25M	-	33.56	7.16	21.49	-	37.29	9.92	22.76
PASS (Oved and Levy, 2021)	440M	-	37.43	8.02	23.34	-	36.91	8.12	23.09
Full (100%)	400M	17.87	37.22	9.17	23.51	12.87	37.40	10.27	23.76
Full (100%) + L1o	400M	16.90	37.67	10.28	24.32	12.40	36.79	11.07	25.03
AdaSum (0.6%)	2.6M	13.45	38.49	9.84	24.37	11.94	37.55	10.11	24.08
AdaSum (0.6%) + L1o	2.6M	12.06	38.94	10.63	24.95	11.23	37.78	11.31	24.04
AdaSum (5%)	21.3M	16.30	38.15	9.18	23.17	12.50	38.12	10.89	24.11
AdaSum (5%) + L1o	21.3M	**12.03**	**39.78**	**10.80**	**25.55**	**11.11**	**38.82**	**11.75**	**25.14**

Test Results - Query-based Summarization



	R1	R2	RL	AR
Full (100%) + Q*	40.52	10.96	25.06	59.84
Full (100%) + L1O + Q*	42.65	11.53	26.82	96.39
AdaQSum (5%)*	41.04	11.08	25.46	60.64
AdaQSum (5%) + L1O*	43.84	13.41	27.31	97.19
AdaQSum (5%)	38.58	10.10	24.19	69.14
AdaQSum (5%) + L1O	38.53	10.52	26.06	98.78

Test Results - Comparison of the Query-based and Generic Summarizers



		R1	R2	RL	unique 1-gram (%)	unique 2-gram (%)
Amazon	AdaSum (5%) + L1O	**39.78**	**10.80**	25.55	67.72	80.83
	AdaQSum (5%) + L1O	38.53	10.52	**26.06**	**69.38**	**82.57**
Yelp	AdaSum (5%) + L1O	**38.82**	**11.75**	**25.14**	62.26	76.55
	AdaQSum (5%) + L1O	36.79	10.06	23.99	**65.74**	**79.88**

CONCLUSIONS

- The proposed method of using adapters and self-supervised pre-training for opinion summarization improves summary quality over standard fine-tuning.

- The self-supervised adapter pre-training method is effective in storing in-domain knowledge and reducing semantic mistakes in generated summaries.

- The proposed method is robust to overfitting on small datasets, which is important in opinion summarization where large annotated datasets are not available.

- The proposed method also addresses the challenge of summary personalization by conditioning on aspect keyword queries, resulting in better-organized summary content reflected in improved coherence and fewer redundancies.
- The experiments on the Amazon and Yelp datasets show that the proposed method outperforms the state-of-the-art methods in terms of ROUGE-L scores.
- The proposed method is efficient and requires only a few annotated samples for fine-tuning, making it suitable for real-world applications.

THANK YOU…