NAACL 2022, Seattle, USA

Efficient Few-Shot Fine-Tuning for Opinion Summarization

Arthur Bražinskas, Ramesh Nallapati
Mohit Bansal, Markus Dreyer

University of Edinburgh
Scotland

Introduction

Motivation

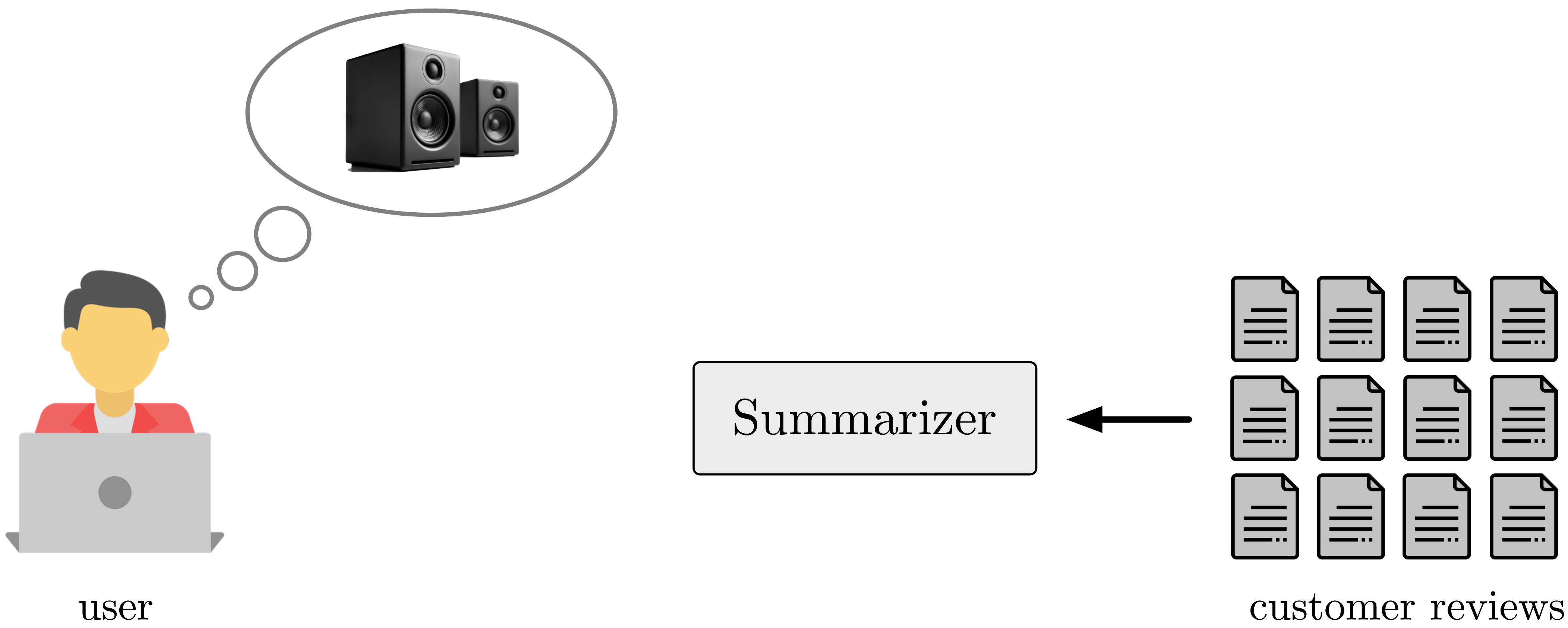
- Users often purchase products **online** (e.g., from Amazon and Google Shopping)
- Seek **opinions** of other **users** expressed in **reviews**
- Use this information for **better purchasing decisions**
- Some products have hundreds or even thousands of reviews — **time consuming to read**

Opinion summarization

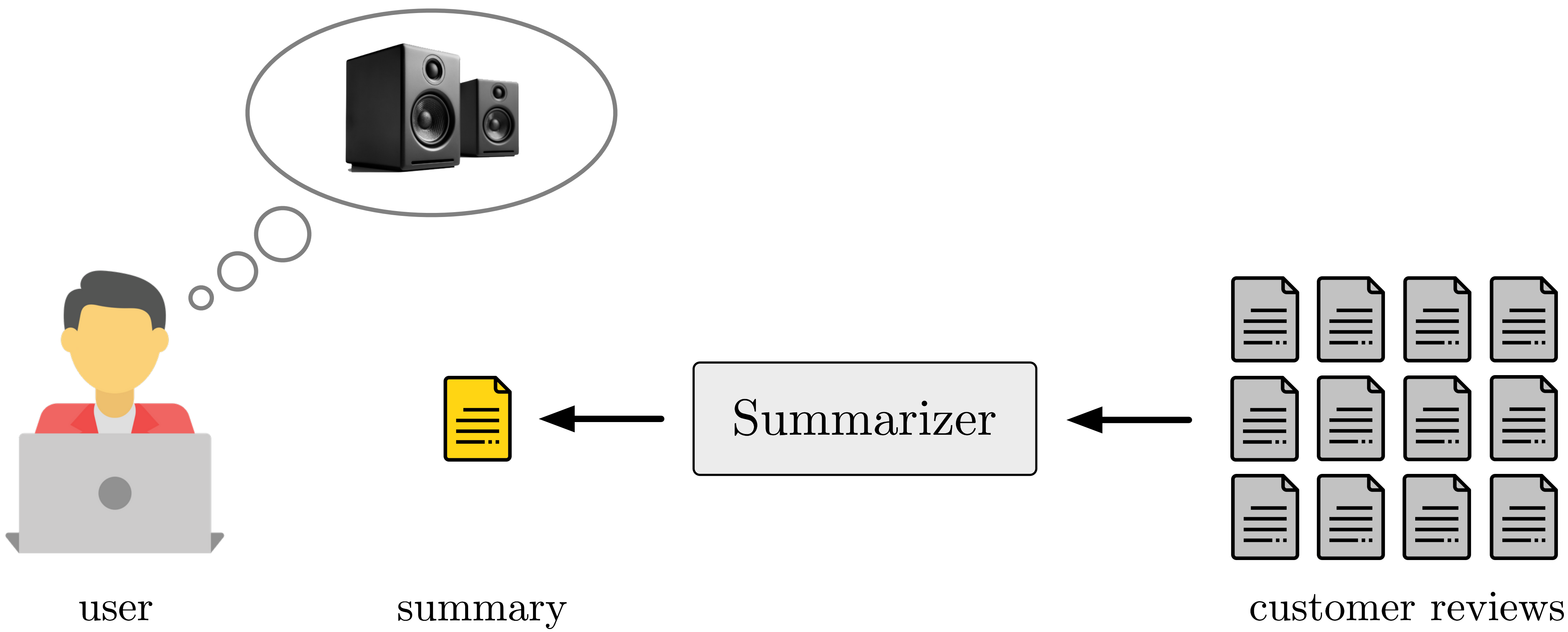
user

customer reviews

Opinion summarization



Opinion summarization



Annotated data scarcity

- Annotated datasets for learning are **scarce**
- Summary writing is **expensive** — many reviews to read
- Most datasets have less than **100** reviews-summary pairs for fine-tuning

Standard fine-tuning

- Standard fine-tuning of PLMs leads to **rapid overfitting**
- Also PLMs are:
  - Pre-trained on **generic text corpora**
  - Often not accustomed to **in-domain specifics**:
    - E.g., product features, aspects, and usage
- Not possible to learn a wide range of product specifics from **a handful of summaries**

Semantic mistakes

This dead on arrival battery is of good quality and holds a charge well. It is easy to install and is a great value for the money. However, it may not hold a charge as advertised due to the plastic case bulging. Overall, this product is highly recommended.

Semantic mistakes

This dead on arrival battery is of good quality and holds a charge well. It is easy to install and is a great value for the money. However, it may not hold a charge as advertised due to the plastic case bulging. Overall, this product is highly recommended.

In this work

- Efficient in-domain knowledge injection via **self-supervised pre-training**
- Reduce **semantic mistakes** in generated summaries
- State-of-the-art results in automatic and human evaluation

Approach

In-domain knowledge

- In-domain knowledge can be learned from **unannotated customer reviews**
- However, further pre-training of a PLM (100% parameters):
  - **Computationally** and **memory inefficient** (Mahabadi et al., 2021)
  - Need a separate copy of the model for each domain (e.g., Yelp, Amazon, and IMDB)
  - Catastrophic forgetting (Chen et al., 2020; Yu et al., 2021)

Adapters

- Use **adapters** (Houlsby et al., 2019)
- Small modules — a few percent of PLM's params
- Inserted into Transformer layers
- PLM is **frozen** while adapters are **optimized**

Self-supervised pre-training

- We use a **self-supervised objective** to learn in-domain knowledge
- Predict a review **conditioned** on the **other reviews** of a product — **leave-one-out** (Bražinskas et al., 2020)

Self-supervised pre-training

Great Italian restaurant with authentic food and great service! Recommend!

review 1

We ordered pasta, and it was very tasty. Would recommend this place to anyone.

review 2

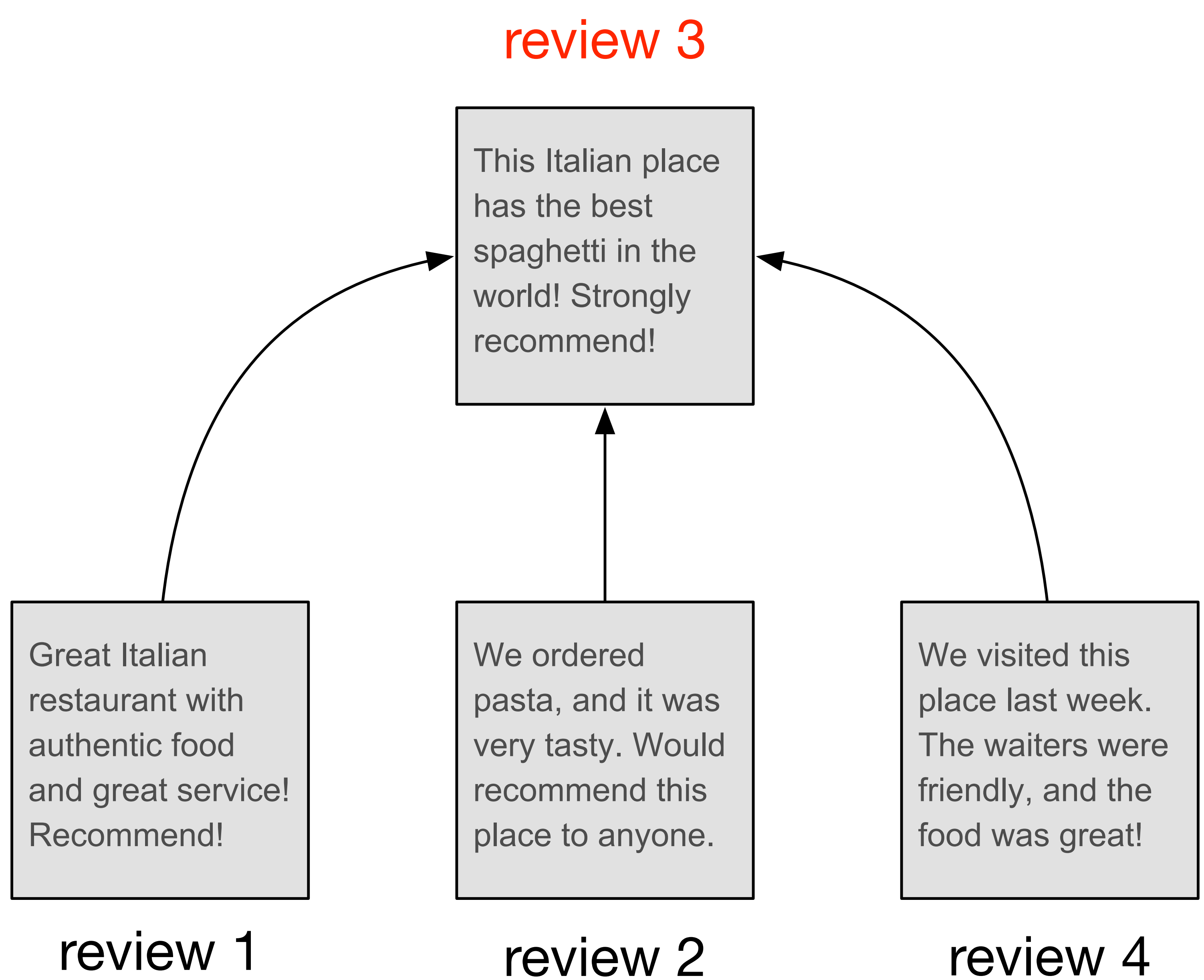
This Italian place has the best spaghetti in the world! Strongly recommend!

review 3

We visited this place last week. The waiters were friendly, and the food was great!

review 4

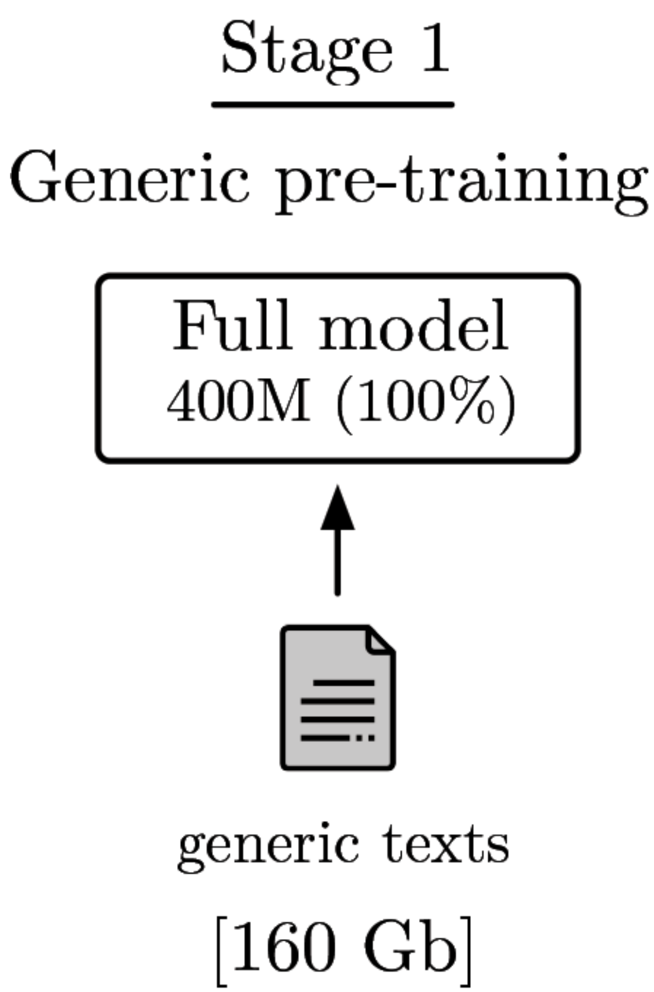
Self-supervised pre-training



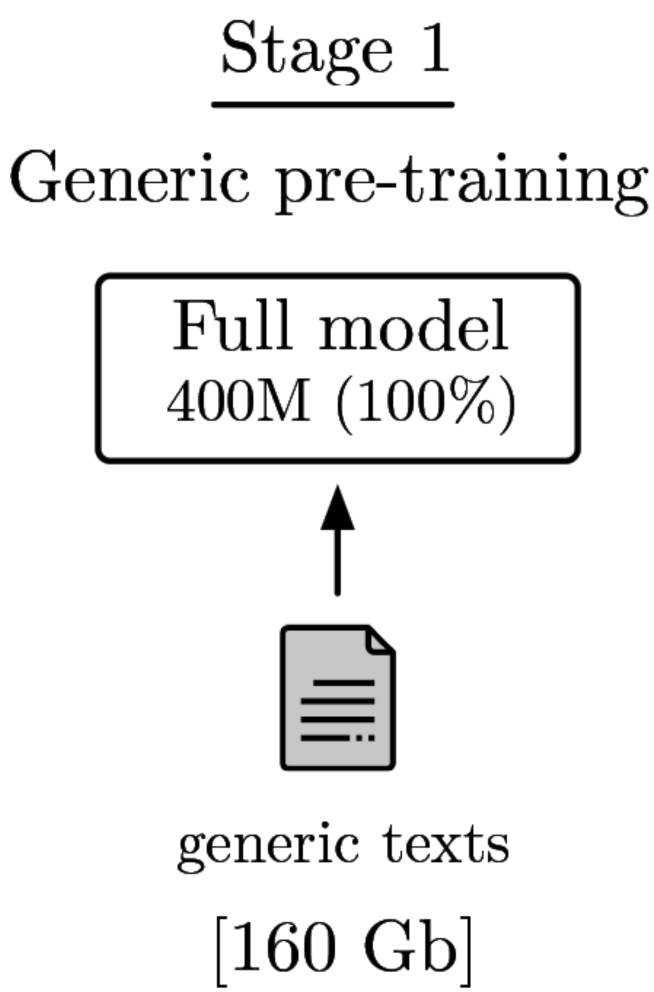
Fine-tuning

- Fine-tune the **pre-trained adapters** on a handful gold samples
- Reviews-summary pairs

Workflow

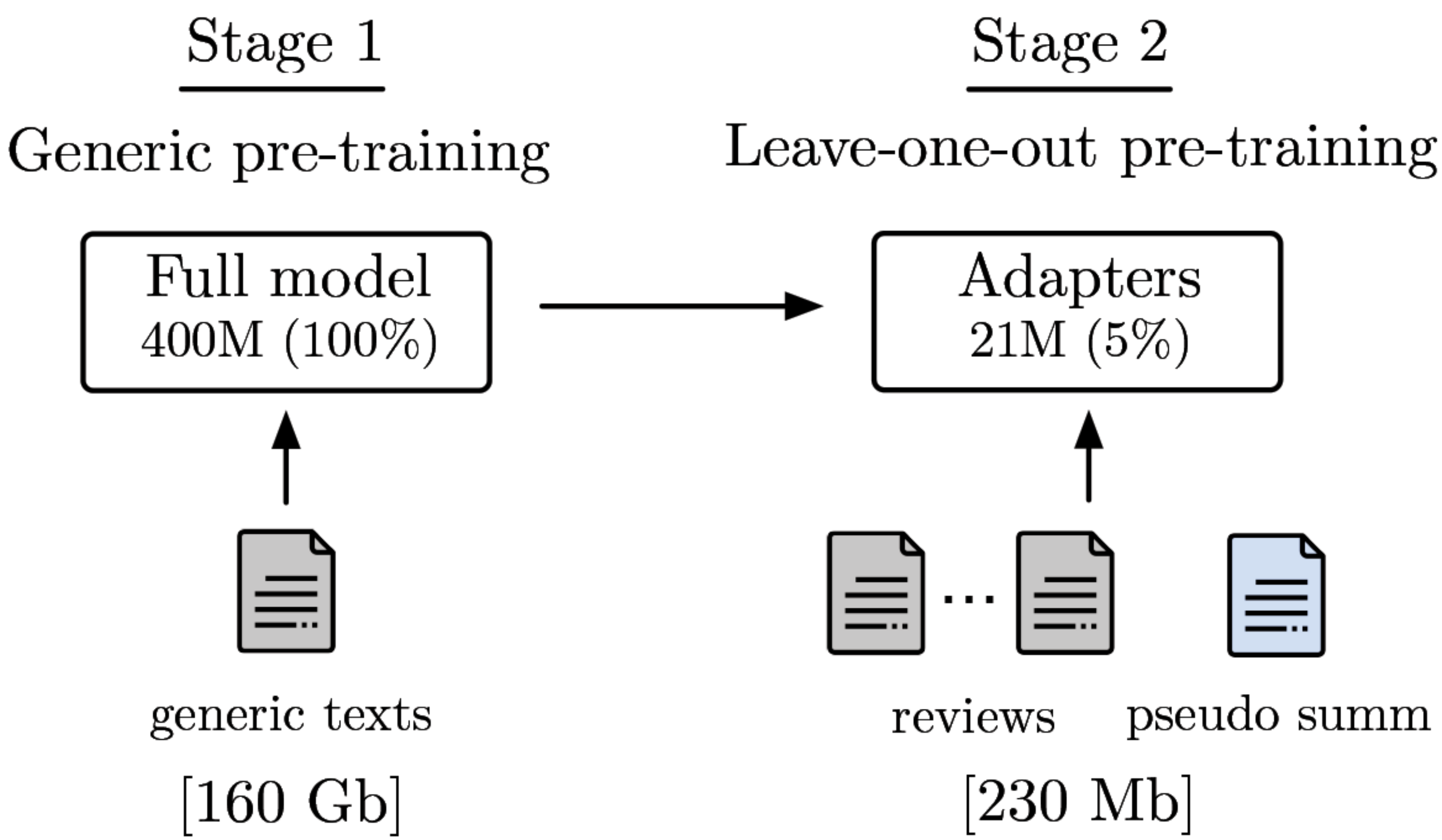

Stage 1
Generic pre-training
Full model
400M (100%)
generic texts
[160 Gb]

Workflow

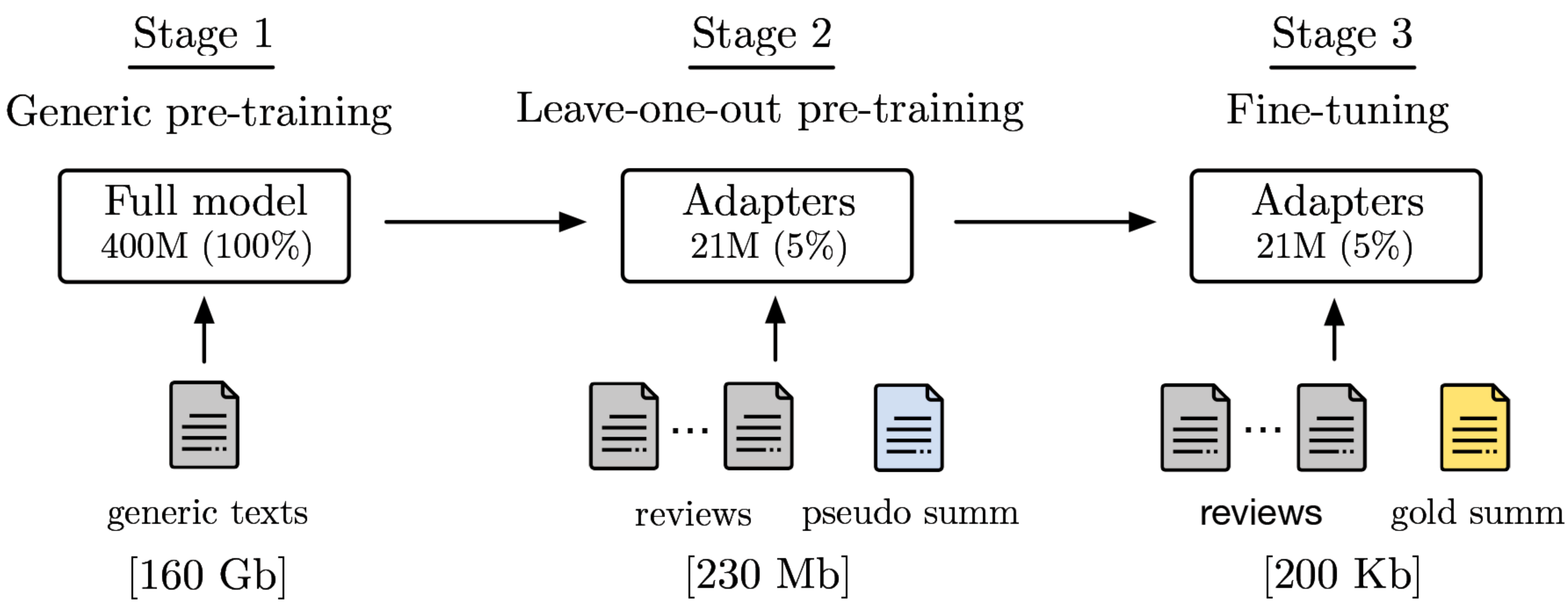


We use BART (Lewis et al., 2020)

Workflow


Stage 1
Generic pre-training
Full model
400M (100%)
generic texts
[160 Gb]
Stage 2
Leave-one-out pre-training
Adapters
21M (5%)
reviews
pseudo summ
[230 Mb]

Workflow


Stage 1
Generic pre-training
Full model
400M (100%)
generic texts
[160 Gb]
Stage 2
Leave-one-out pre-training
Adapters
21M (5%)
reviews
pseudo summ
[230 Mb]
Stage 3
Fine-tuning
Adapters
21M (5%)
reviews
gold summ
[200 Kb]

Experimental Setup

Scope

- Will present results on **Amazon**
- **Results** on Yelp can be found in the paper

Data

	Pre-training data (He and McAuley, 2016)	Fine-tuning data (Bražinskas et al., 2020)
Train	70,144	84
Valid	7,900	36
Test	-	60

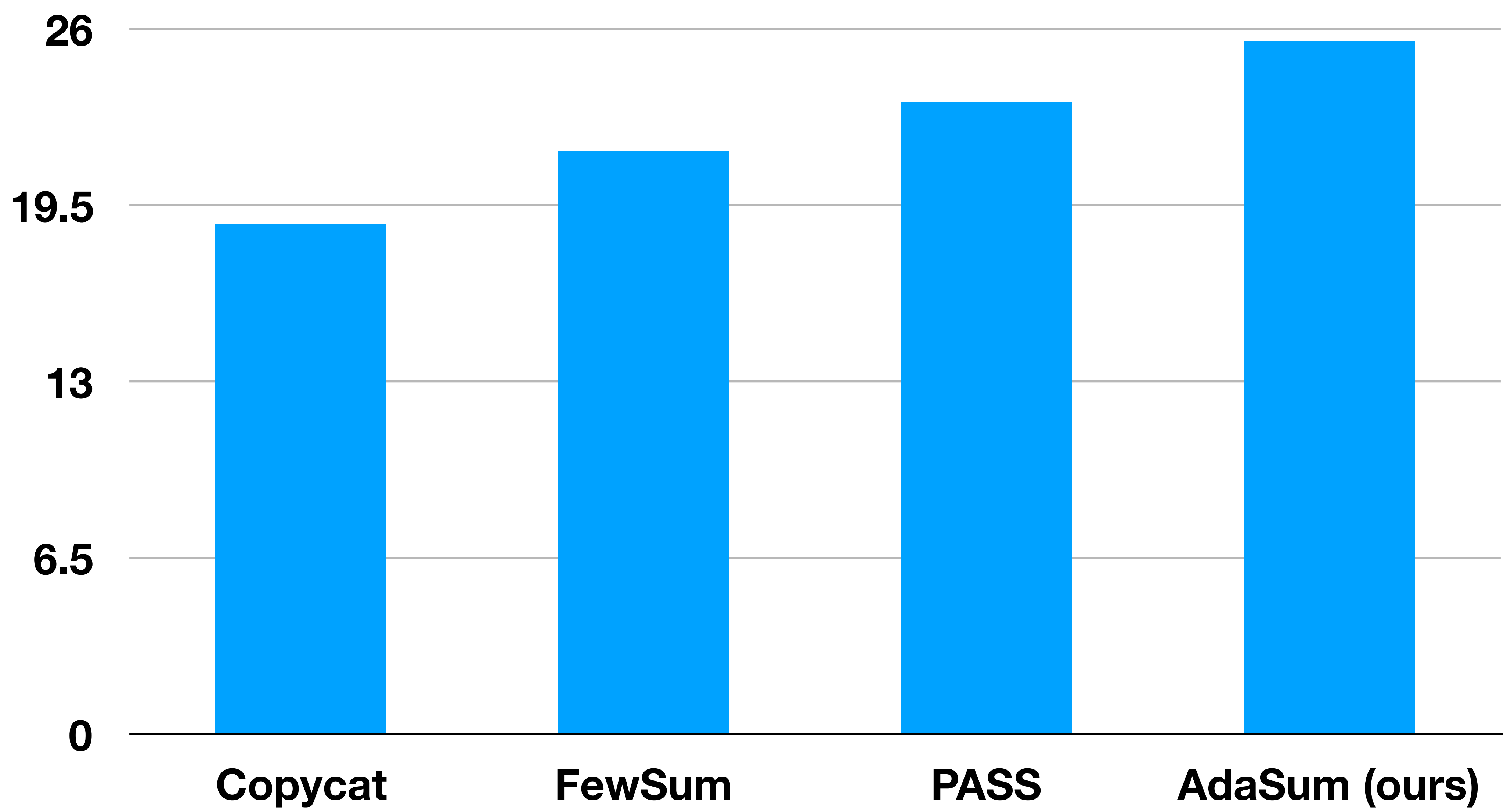
amazon

Baselines

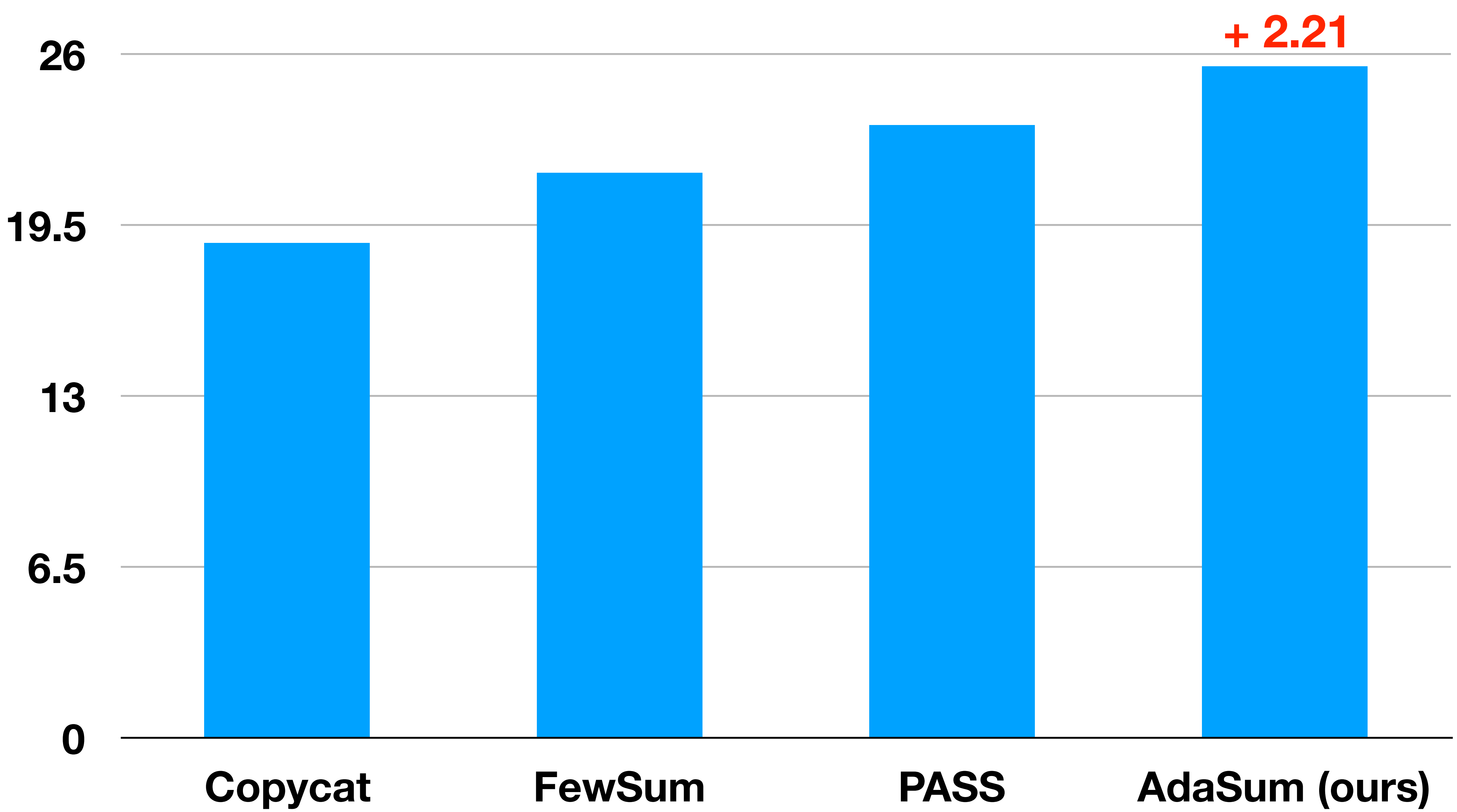
- **Copycat** (Bražinskas et al., 2020): unsupervised model
- **FewSum** (Bražinskas et al., 2020): few-shot model
- **PASS** (Oved and Levi, 2022): SOTA few-shot model
- …

Results

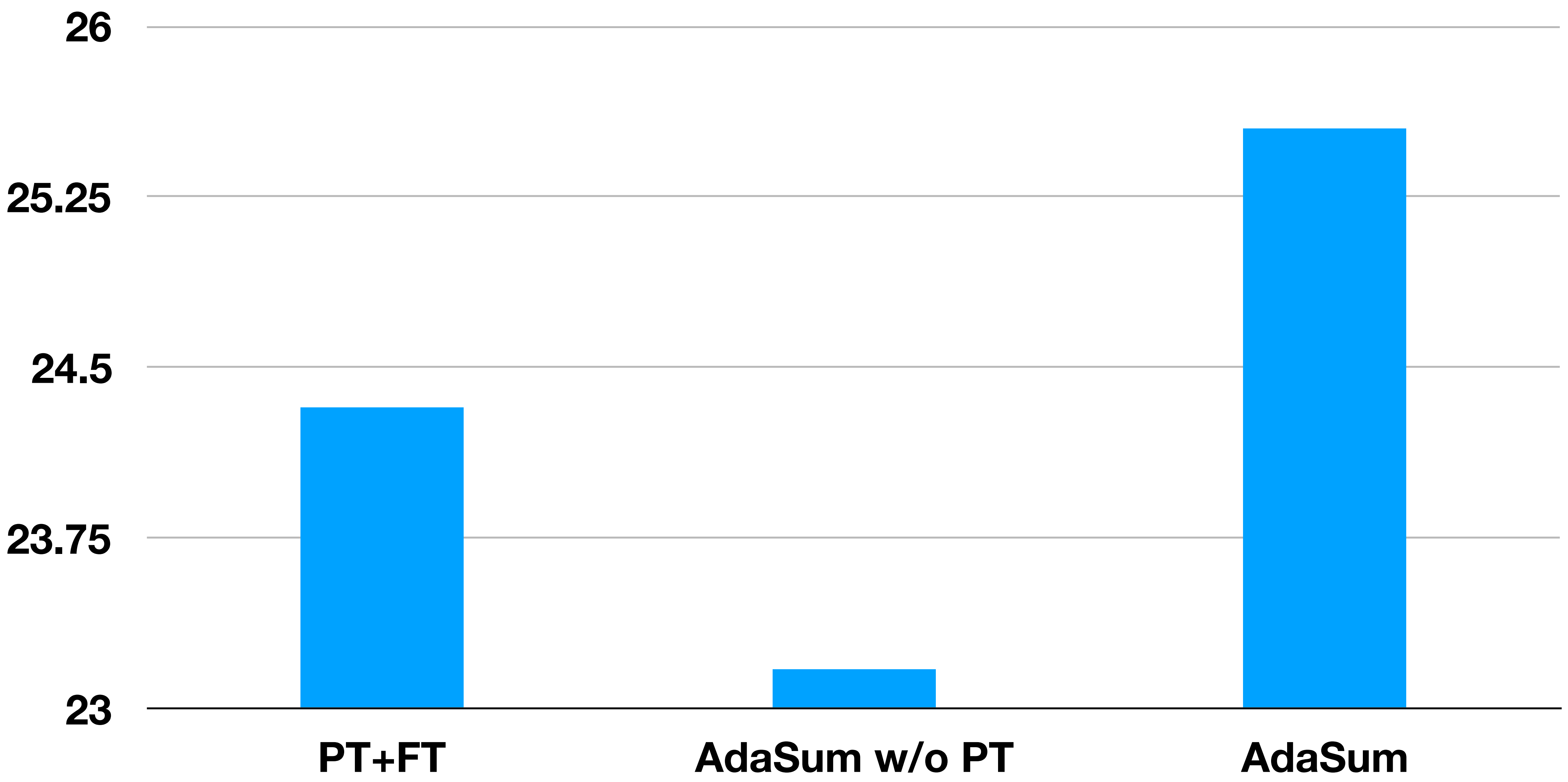
ROUGE-L scores



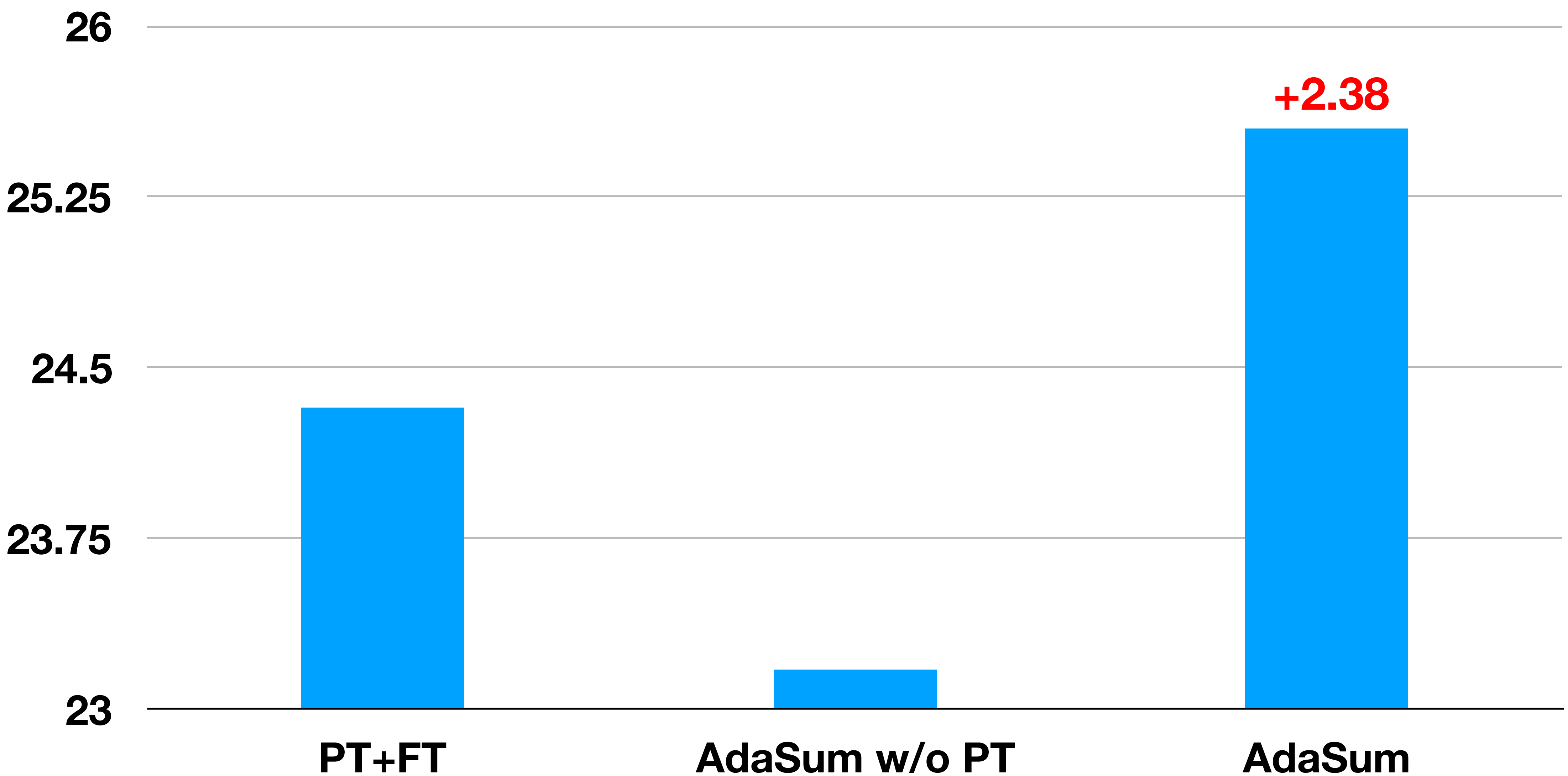
ROUGE-L scores



ROUGE-L scores



ROUGE-L scores



Conclusions

Main contributions

- Proposed in-domain knowledge injection into **adapters** via **self-supervised** pre-training
- Reduce **semantic mistakes** in summaries
- SOTA results
- …

In the paper

- Aspect-based model learned from **a handful of gold samples** after **self-supervised pre-training**
- Detailed analysis: abstractivness, semantic mistakes, etc
- More automatic and human evaluation results