Low-Resource Learning

(30 minutes)

Low-Resource learning

How do we **learn models** from small **annotated** datasets?

Modern low-resource learning: 2-steps

1) **Pre-training** on large collections of **generic corpora**
2) **Fine-tuning** on gold samples (e.g., reviews-summary pairs)

Fine-tuning

- Often performed on **thousands of gold samples** in other branches of summarization
- For example, single- and multi-document news summarization

News summarization annotated datasets

	#Samples	Multi-document input?
CNN/DailyMail (Hermann et al., 2014)	311,971	No

News summarization annotated datasets

	#Samples	Multi-document input?
CNN/DailyMail (Hermann et al., 2014)	311,971	No
XSum (Narayan et al., 2018)	226,183	No

News summarization annotated datasets

	#Samples	Multi-document input?
CNN/DailyMail (Hermann et al., 2014)	311,971	No
XSum (Narayan et al., 2018)	226,183	No
Multi-News (Fabbri et al., 2019)	56,216	Yes
…	…	…

Annotated data in opinion summarization

- Annotators need to read many reviews to write each summary
- This has lead to only a **few datasets** with **small numbers of summaries**

Available datasets

	#Entities	#Summaries	Domain
MeanSum (Chu and Liu, 2019)	200	200	Yelp

Available datasets

	#Entities	#Summaries	Domain
MeanSum (Chu and Liu, 2019)	200	200	Yelp
Copycat (Bražinskas et al., 2020)	60	180	Amazon

Available datasets

	#Entities	#Summaries	Domain
MeanSum (Chu and Liu, 2019)	200	200	Yelp
Copycat (Bražinskas et al., 2020)	60	180	Amazon
FewSum (Bražinskas et al., 2020)	60	180	Amazon

Available datasets

	#Entities	#Summaries	Domain
MeanSum (Chu and Liu, 2019)	200	200	Yelp
Copycat (Bražinskas et al., 2020)	60	180	Amazon
FewSum (Bražinskas et al., 2020)	60	180	Amazon
SpaCe (Angelidis et al., 2021)	50	1,050	TripAdvisor

Challenges

Challenges

- Large neural models tend to **rapidly overfit** on small datasets
- There is a number of `dimensions' **challenging to learn**
  - Summary characteristics (e.g., the style of writing)
  - In-domain specifics (product aspects and details)
  - Content structure
  - Personalized summarization (e.g., aspect-based summarization)

Summary characteristics

- **Style-of-writing**: don’t want summaries written in **the style of reviews**
- **Summary informativeness**: don’t want generic content
- …

Summary characteristics: example

This is my second pair of Reebok running shoes and they are the best running shoes I have ever owned. They are lightweight, comfortable, and provide great support for my feet.

Summary characteristics: example

review writing style and uninformative

This is my second pair of Reebok running shoes and they are the best running shoes I have ever owned. They are lightweight, comfortable, and provide great support for my feet.

Summary characteristics: example

This is my second pair of Reebok running shoes and they are the best running shoes I have ever owned. They are lightweight, comfortable, and provide great support for my feet.

review writing style

In-domain specifics

- Online shops sell a multitude of different products (electronics, health, etc)
- Products can have different specifics:
  - Features / Aspects
  - Utilization / Usage
- It’s not possible to learn **a wide range of in-domain specifics** from a **handful of gold samples**

Semantic mistakes

- Results in subtle **semantic mistakes** in generated summaries
- Hard to detect using **automatic evaluation metrics**

Semantic mistakes: example

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: example

semantic mistake

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.

Content structure

- Reviewers often disagree on the pros and cons of a given product
- Summarizers sometimes yield **inconsistent** and **self-contradicting** summaries (Oved and Levi, 2021)

Self-contradictions: example

These running shoes are great! They fit true to size and are very comfortable to run around in. They are light weight and have great support. They run a little on the narrow side, so make sure to order a half size larger than normal.

Self-contradictions: example

These running shoes are great! They fit true to size and are very comfortable to run around in. They are light weight and have great support. They run a little on the narrow side, so make sure to order a half size larger than normal.

Personalized summarization

- Users often have **particular preferences** when search for products
- These preferences can be used to generate more **`targeted’** summaries reflecting these preferences

Personalized summarization

Preferences can be expressed as:

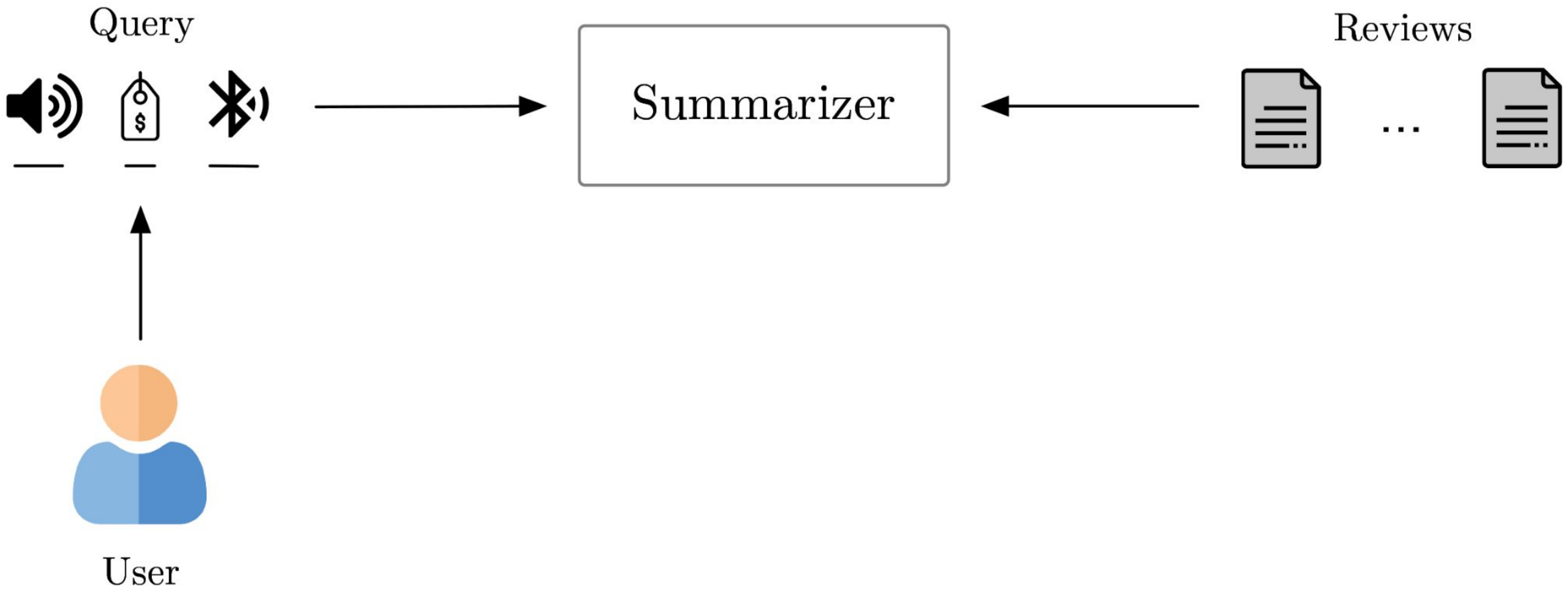
1) Profile information (explicit entries about interests)
2) Aspect queries, e.g., `resolution`, `battery life’, `price’

…

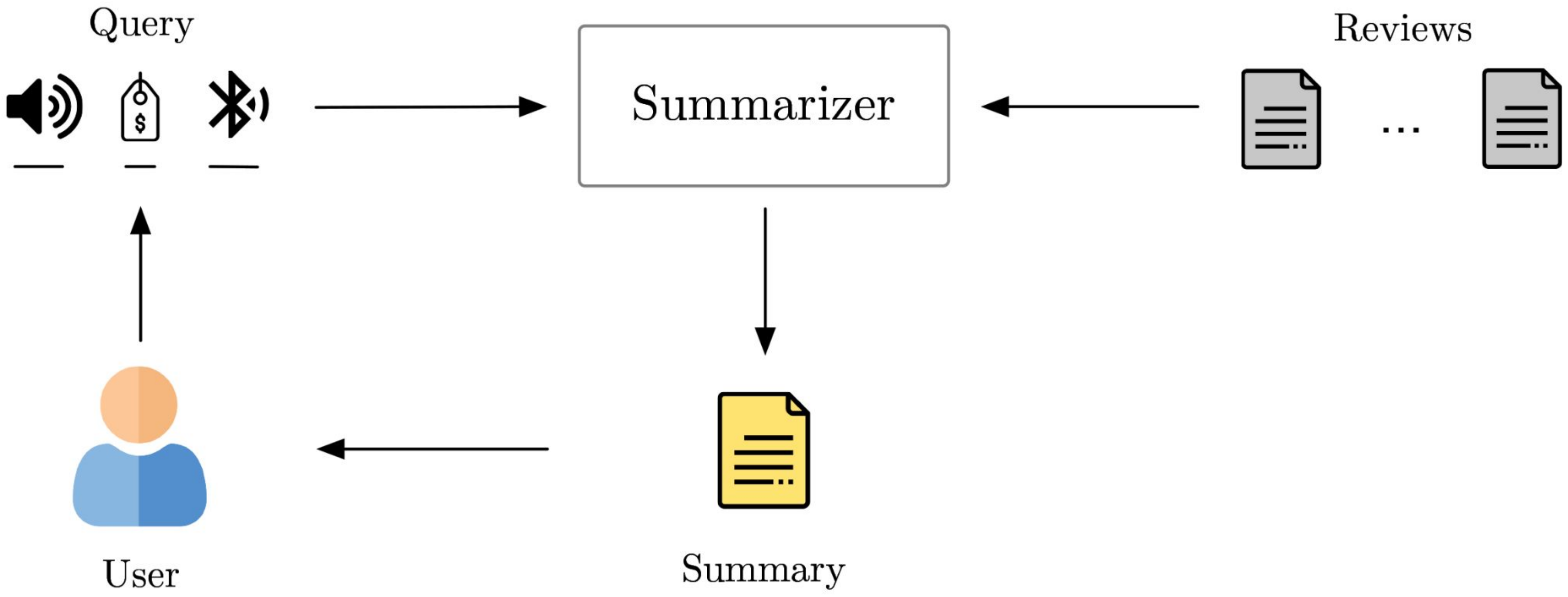
Personalized summarization: illustration

Personalized summarization: illustration

Personalized summarization: illustration



Personalized summarization: illustration



Personalized summarization

This task is also challenging to learn as the model needs to:

1) Rely on the query
2) Generate the summary **corresponding** to the query

Also, in test time, we can assume that there can be **a wide of range of queries**

Propose solutions

Challenge	Proposed Solution	Main ideas
summary characteristics	**FewSum** (Bražinskas et al., 2020)	explicitly model summary characteristics

Propose solutions

Challenge	Proposed Solution	Main ideas
summary characteristics	**FewSum** (Bražinskas et al., 2020)	explicitly model summary characteristics
in-domain specifics	**AdaSum** (Bražinskas et al., 2022)	learn in-domain specifics from customer reviews

Propose solutions

Challenge	Proposed Solution	Main ideas
summary characteristics	**FewSum** (Bražinskas et al., 2020)	explicitly model summary characteristics
in-domain specifics	**AdaSum** (Bražinskas et al., 2022)	learn in-domain specifics from customer reviews
content structure	**PASS** (Oved and Levi, 2021)	1) generate multiple summaries 2) rank summaries by coherence

Propose solutions

Challenge	Proposed Solution	Main ideas
summary characteristics	**FewSum** (Bražinskas et al., 2020)	explicitly model summary characteristics
in-domain specifics	**AdaSum** (Bražinskas et al., 2022)	learn in-domain specifics from customer reviews
content structure	**PASS** (Oved and Levi, 2021)	1) generate multiple summaries 2) rank summaries by coherence
personalization (aspect queries)	**AdaQSum** (Bražinskas et al., 2022)	1) automatically create aspect queries 2) learn the task from customer reviews

Challenge: Summary Characteristics

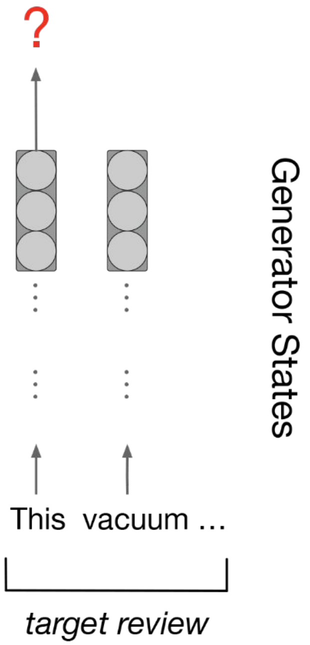
FewSum

- **FewSum**: **few-shot learning model** (Bražinskas et al., EMNLP 2020)
- **Summary characteristics** are **modelled** explicitly
- Utilizes a **handful of human-written summaries** for fine-tuning

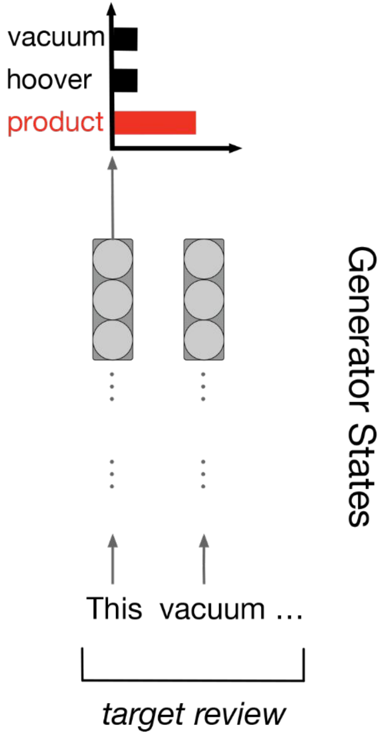
Architecture

- Similar to Copycat: conditional language model (CLM)
- Encoder-decoder architecture (Transformers without pre-initialization)
- In-domain **pre-training** on a collection of customer reviews via **leave-one-out**

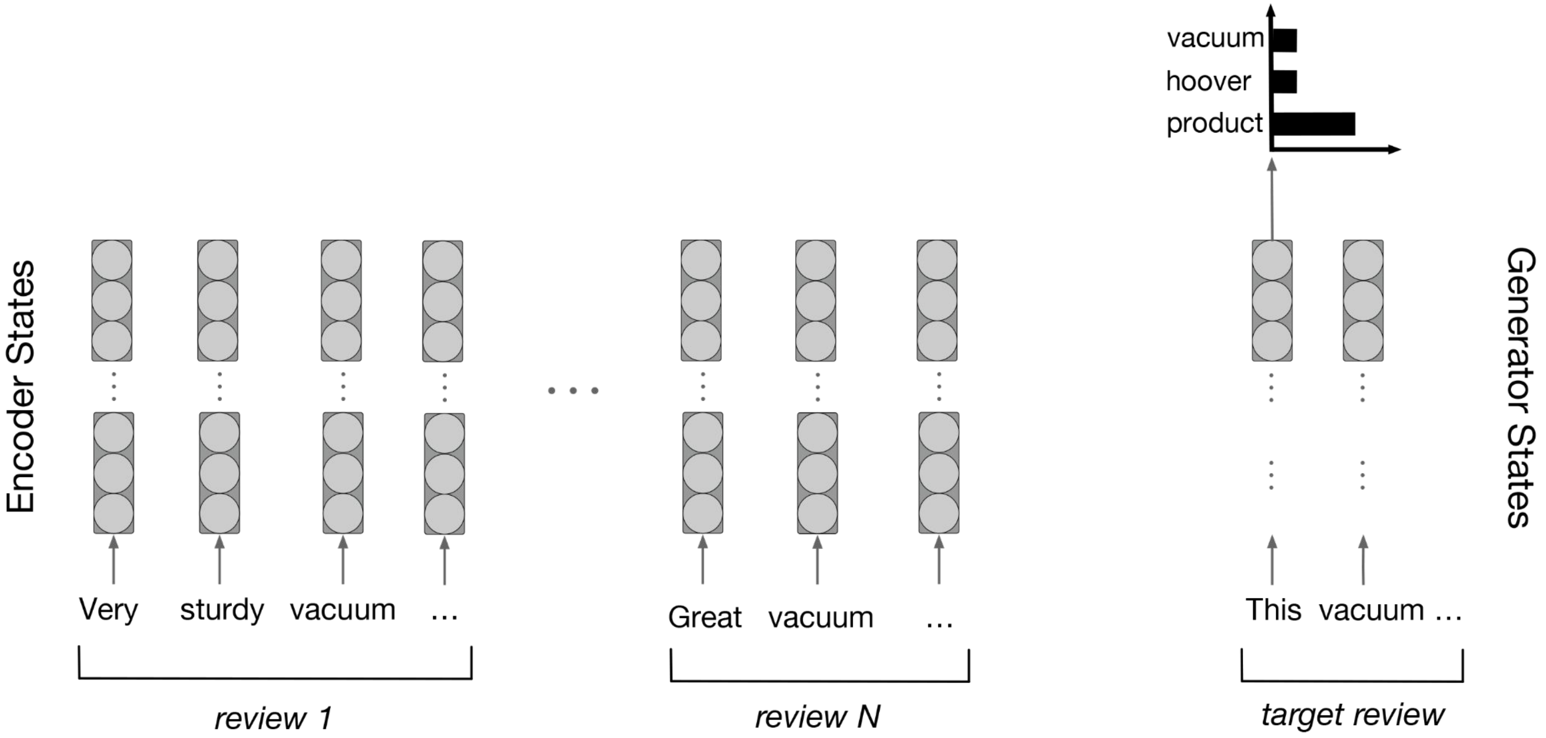
Leave-one-out



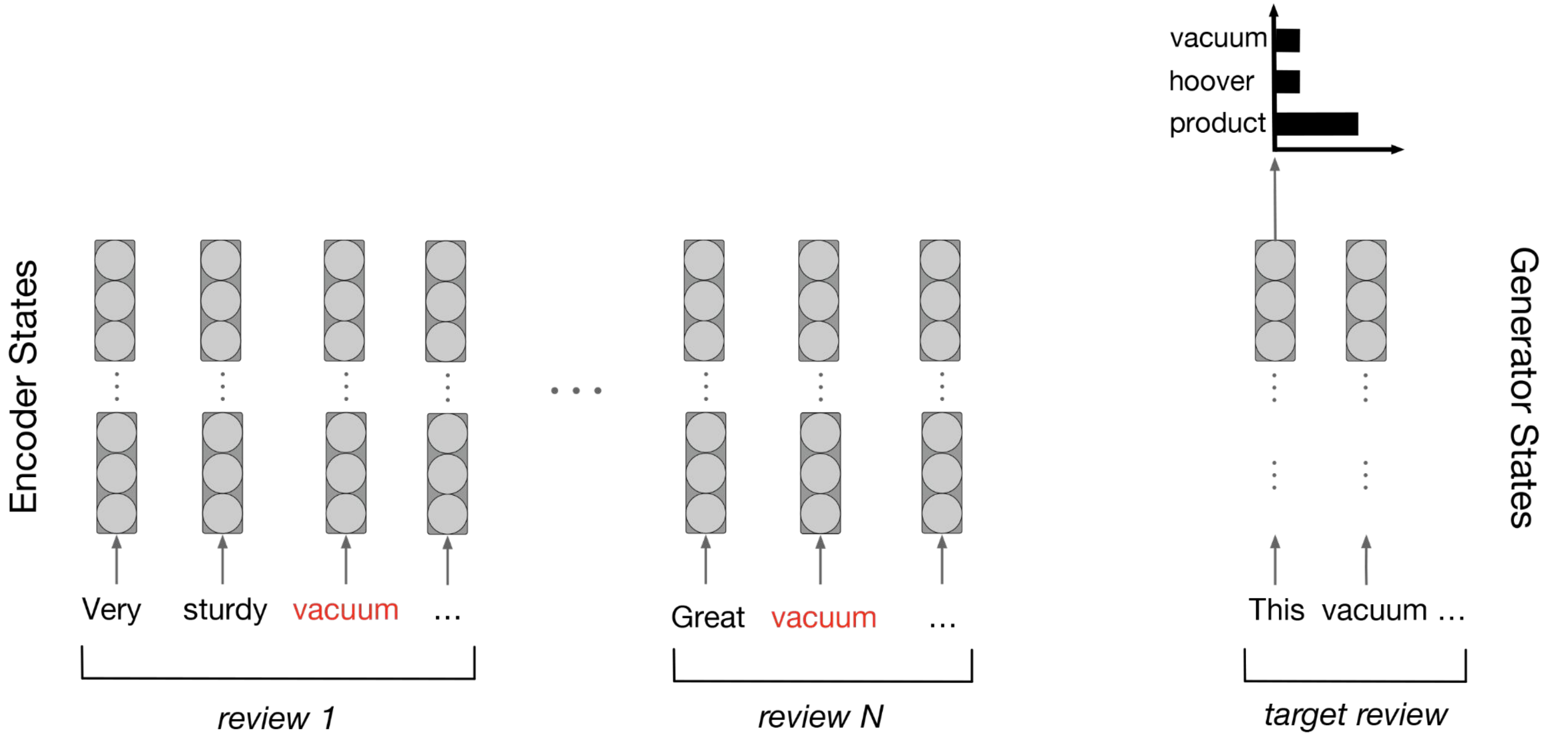
Leave-one-out



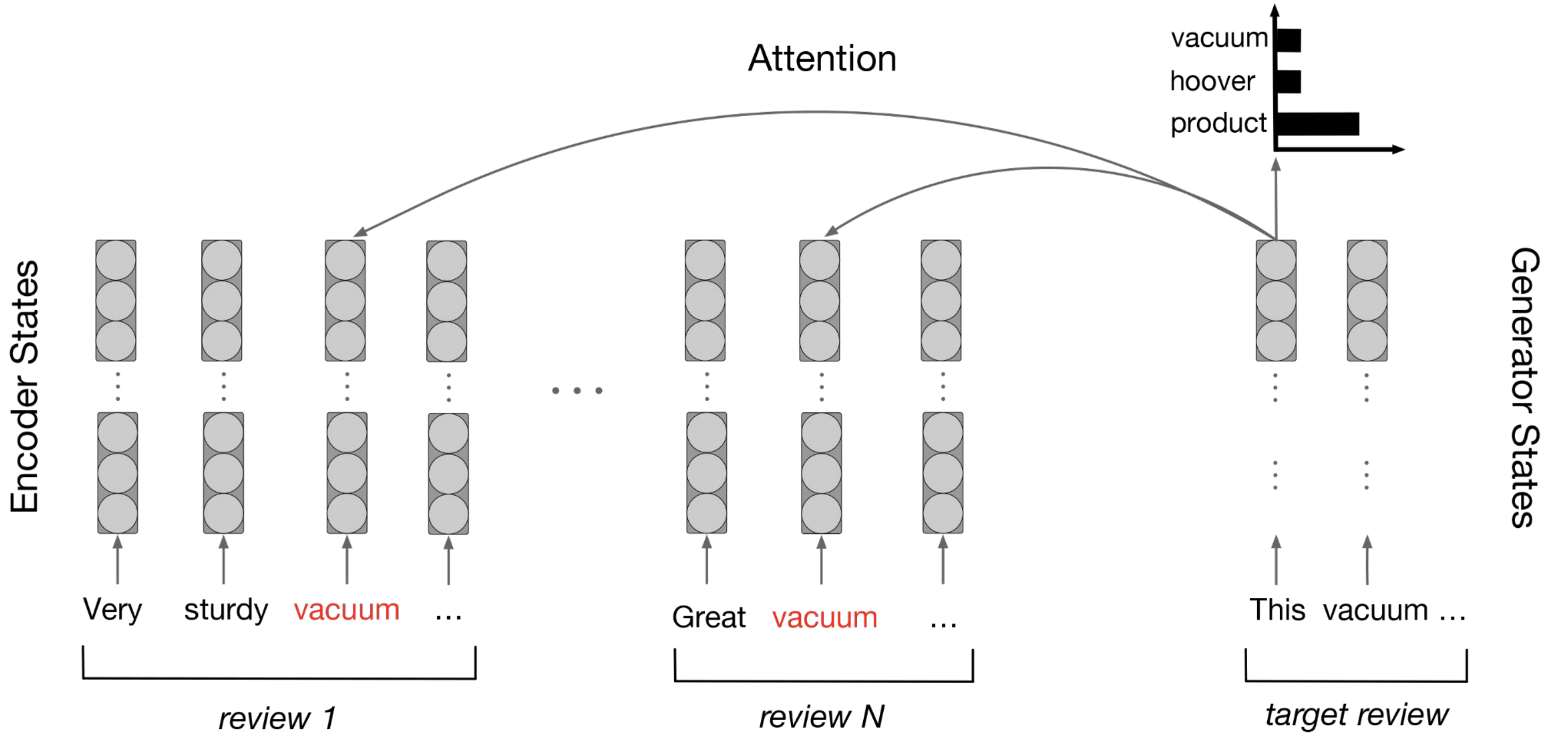
Leave-one-out



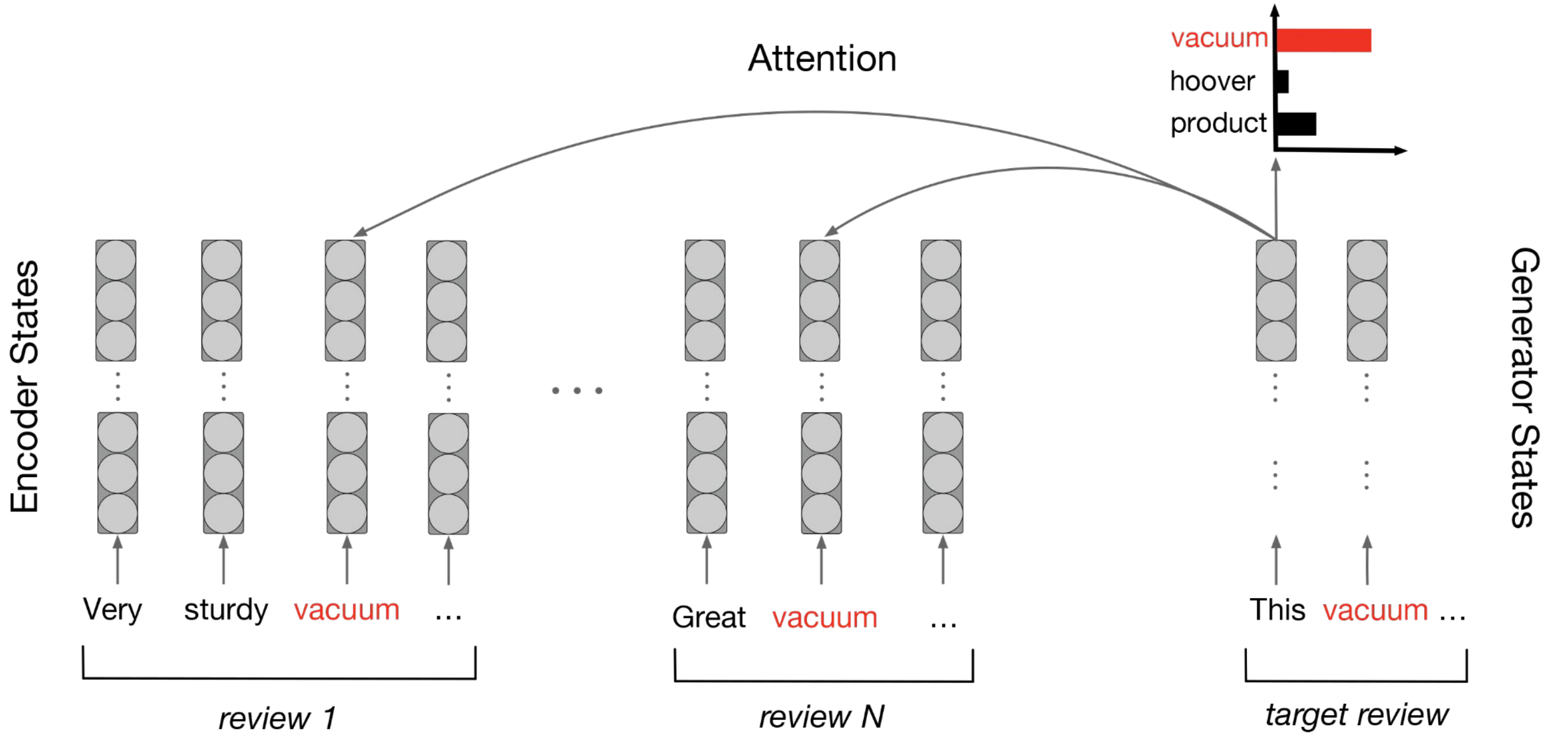
Leave-one-out



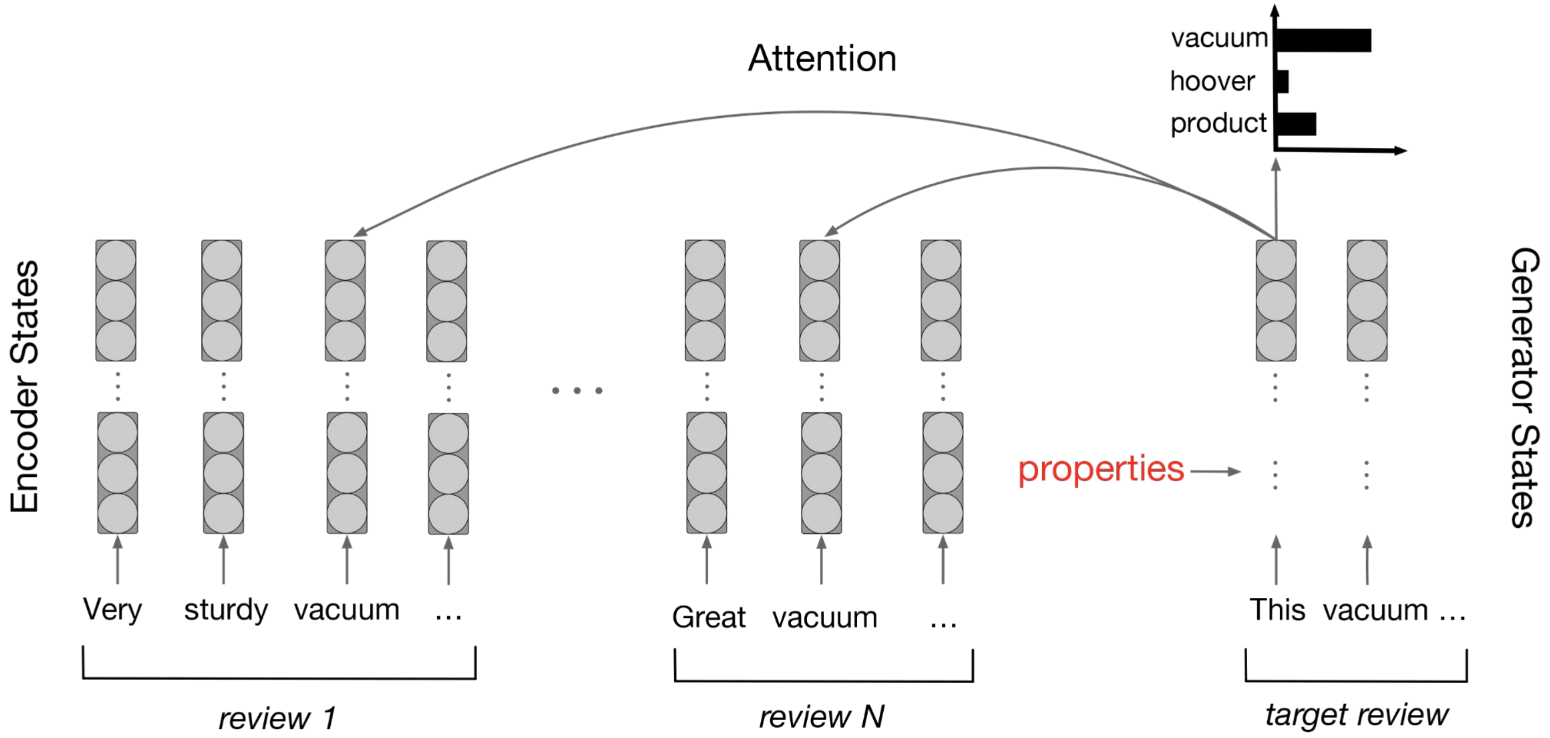
Leave-one-out



Leave-one-out



Leave-one-out



Properties

Observation:

- Some reviews are more like summaries
- Some are less

Review 1

Varys

When I first got diabetes I got this. It has a lot of what we need. But later I have switched to another brand.

Review 1

Varys

Not Informative
+ informal writing style

When I first got diabetes I got this. It has a lot of what we need. But later I have switched to another brand.

Review 1

Varys

★★★☆☆

When I first got diabetes I got this. It has a lot of what we need. But later I have switched to another brand.

Review 2

Jon Snow

These capsules are a natural alternative to other over-the-counter medications. They are easy to swallow and have a great taste. Overall, great value for money.

Review 2

Jon Snow

☆☆☆☆☆

These capsules are a natural alternative to other over-the-counter medications. They are easy to swallow and have a great taste. Overall, great value for money.

Review 2

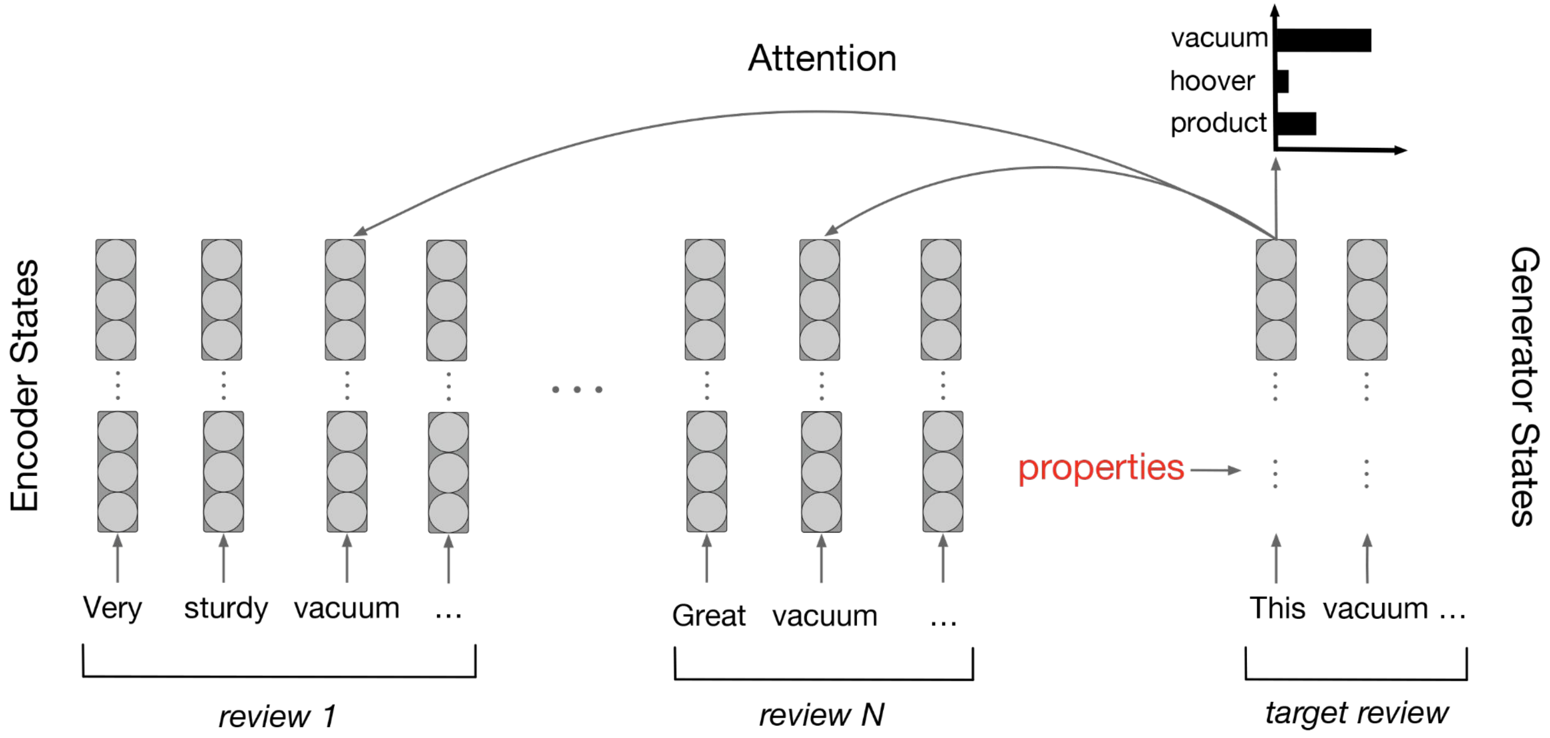
Jon Snow

These capsules are a natural alternative to other over-the-counter medications. They are easy to swallow and have a great taste. Overall, great value for money.

Properties

- Summary characteristics are modelled in terms of **properties**
- `Tell' the decoder how **`summary-like'** a target review is
- Similar to constrained codes (MacKay, 2003)

Properties



Attention
vacuum
hoover
product
Encoder States
Generator States
properties
Very sturdy vacuum ...
Great vacuum ...
This vacuum ...
review 1
review N
target review

Property types

Type	Reviews	Summaries	Implementation
Information coverage	Uncommon	Common	ROUGE scores

Property types

Type	Reviews	Summaries	Implementation
Information coverage	Uncommon	Common	ROUGE scores
Writing style	Informal	Formal	Pronoun counts

Property types

Type	Reviews	Summaries	Implementation
Information coverage	Uncommon	Common	ROUGE scores
Writing style	Informal	Formal	Pronoun counts
…	…	…	…

Oracle

- For training, an oracle function is used
- Automatically computes properties based on:
    - target review
    - source reviews

Oracle

$$q(\text{target review}, \text{source reviews})$$

Plug-in network

- At test time, want to generate **summaries**
- Have access only to **source reviews**
- Can’t use the **oracle**
- Might **not know** what **property assignments** are needed to generate summaries
- Replace the **oracle** by a **trainable neural network**

Plug-in network

- Using a **handful** of summaries (< 30 products)
- Fine-tune the **plug-in network**
- Learns what property assignments lead to **generation of summaries**

Workflow

1) Pre-training on customer reviews
2) Fine-tuning of the plug-in network on annotated summaries
3) Summary generation

Pre-training

- Use a large corpus of reviews
- **Leave-one-out** objective
- **Oracle** computes **property assignments** for **target reviews**

Fine-tuning

- Replace the oracle by the **plug-in network**
- It relies only on input/source reviews
- Fine-tune it on a **handful** of **human-written summaries**

Summary generation

- Use the **plug-in network** to yield properties
- Generate summaries

Summary example

FewSum	These running shoes are great! They fit true to size and are very comfortable to run around in. They are light weight and have great support. They run a little on the narrow side, so make sure to order a half size larger than normal.
Reviews	perfect fit for me ... supply the support that I need ... are flexible and comfortable ... || ... It is very comfortable ... I enjoy wearing them running ... || ... running shoes ... felt great right out of the box ... They run true to size ... || ... my feet and feel like a dream ... Totally light weight ... || ... shoes run small ... fit more true to size ... fit is great! ... supports my arch very well ... || ... They are lightweight... usually wear a size women's 10 ... ordered a 10.5 and the fit is great!

FewSum outputs

- Better capture **expected summary characteristics**
- Are more preferred by people to alternative models (e.g., Copycat) in terms of
  - Fluency
  - Coherence
  - Non-redundancy
  - Informativeness
  - Sentiment alignment

Challenge: Learning In-domain Specifics

Generic PLMs

- PLMs have millions of parameters (e.g., ~400M in BART)
- **Rapidly overfit** during the fine-tuning on a handful of gold samples (<100)
- Rarely accustomed to in-domain specifics
  - What products are about (features and details)
  - How products are used by customers
- Result in **subtle semantic mistakes** in generated summaries after fine-tuning

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.

Proposed solution

- Proposed in (Bražinskas et al., NAACL 2022)
- Learning of **in-domain specifics** from **customer reviews**
- Store the knowledge to **separate modules** – not part of the pre-trained LM
- Reduces **semantic mistakes** and results in SOTA results

Approach

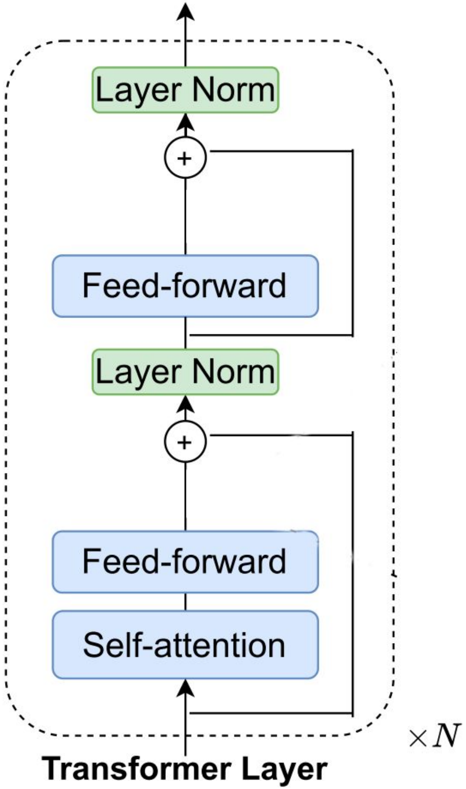
PLM fine-tuning

- In this setting, all parameters are optimized
- Leads to **rapid overfitting** in low-resource settings (He et al., 2021)

Adapters

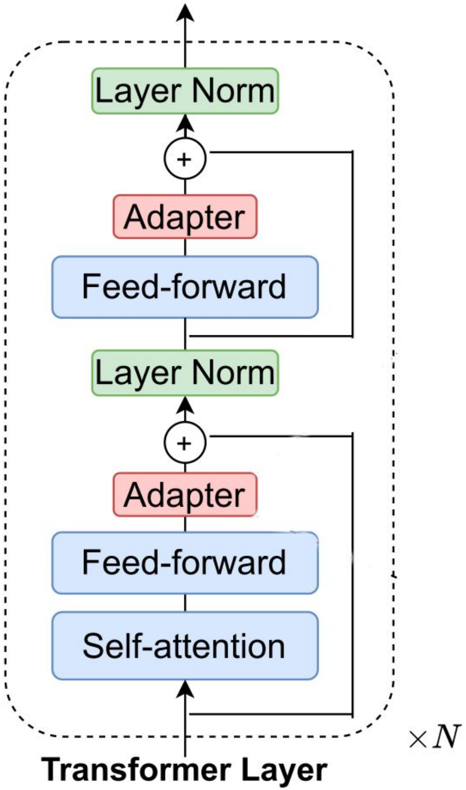
- Use **adapters** (Houlsby et al., 2019)
- Small modules — a few percent of PLM's params
- Inserted into Transformer layers
- PLM is **frozen** while adapters are **optimized**

Adapters

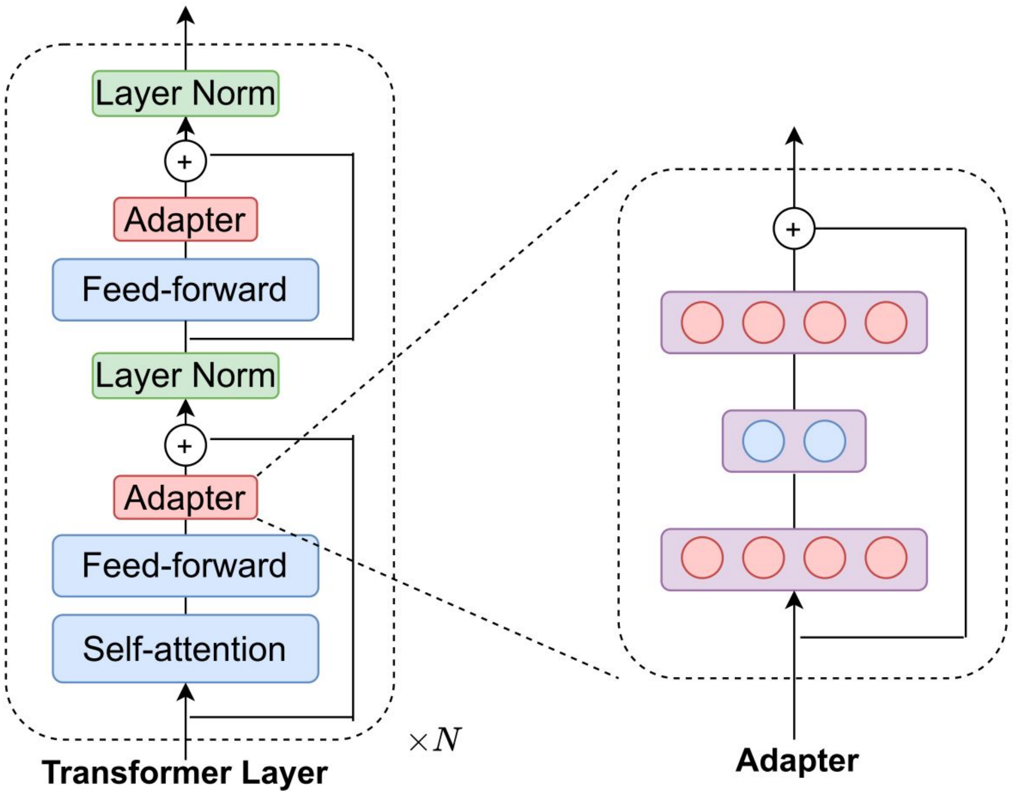


Layer Norm
+
Feed-forward
Layer Norm
+
Feed-forward
Self-attention
$\times N$
Transformer Layer

Adapters


Layer Norm
Adapter
Feed-forward
Layer Norm
Adapter
Feed-forward
Self-attention
$\times N$
Transformer Layer

Adapters


Layer Norm
+
Adapter
Feed-forward
Layer Norm
+
Adapter
Feed-forward
Self-attention
Transformer Layer
$\times N$
+
Adapter

Adapters

$$\hat{h} = f_2(\tanh f_1(h)) + h$$

Adapters

input hidden state

$$\hat{h} = f_2(\tanh f_1(h)) + h$$

Adapters

input hidden state

$$\hat{h} = f_2(\tanh f_1(h)) + h$$

down-projection layer (trainable)

Adapters

up-projection layer (trainable)   input hidden state

$$\hat{h} = f_2(\tanh f_1(h)) + h$$

down-projection layer (trainable)

In-domain knowledge

- In order to **reduce semantic mistakes**, we can learn **in-domain specifics** from **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)

Self-supervised pre-training

- Use a variant of **leave-one-out** to pre-train adapters
- The PLM remains **frozen** during pre-training

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

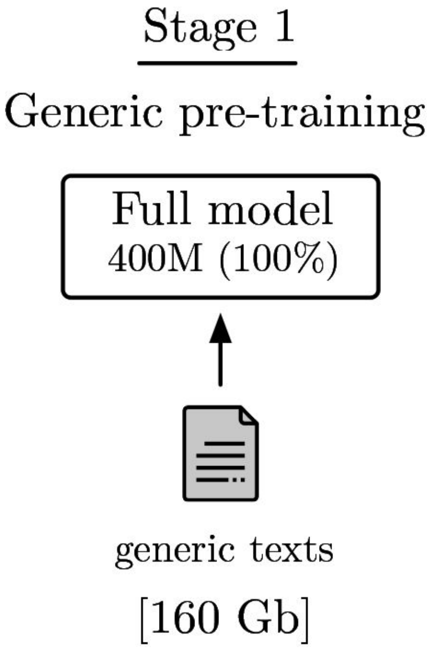
Self-supervised pre-training

- Customer reviews are available in **large quantities** (millions)
- This allows the model to learn a **wide range of in-domains specifics**

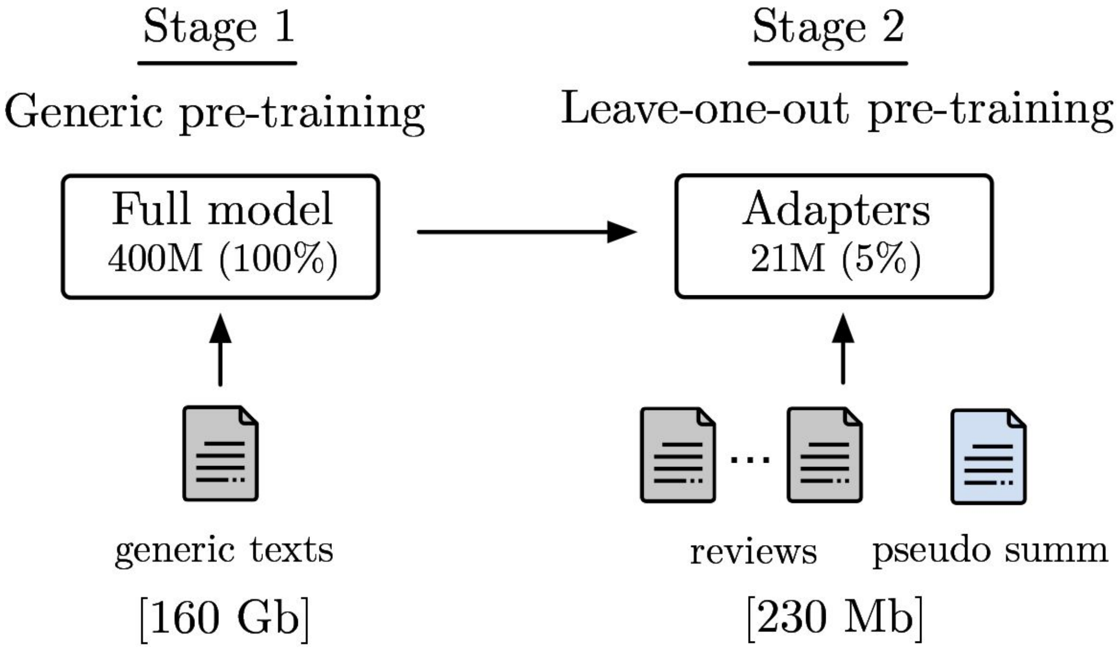
Fine-tuning

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

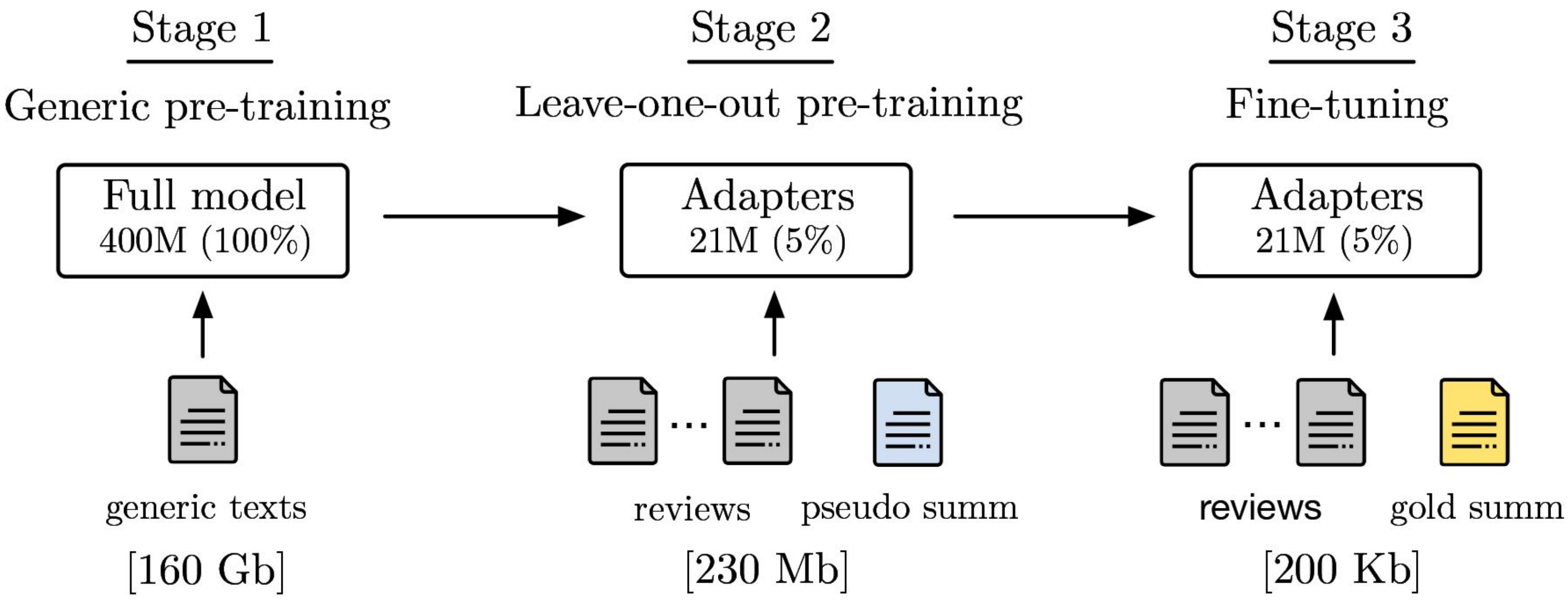
Workflow


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

Workflow


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]

Summary: without pre-training

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.

Summary: without pre-training

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.

Summary: with pre-training

This battery is a great value for the price and works great. It is a good quality battery that can be used to replace a dead battery in an alarm system. The price is great and the quality of the product is good. The shipping was fast and the customer service was excellent.

Summary: with pre-training

This battery is a great value for the price and works great. It is a good quality battery that can be used to replace a dead battery in an alarm system. The price is great and the quality of the product is good. The shipping was fast and the customer service was excellent.

Challenge: Learning Summary Structure

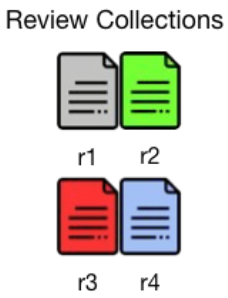
In-coherent summaries

- Reviewers often **disagree** on the pros and cons of a given product
- Summarizers sometimes yield **inconsistent**, **self-contradicting** summaries

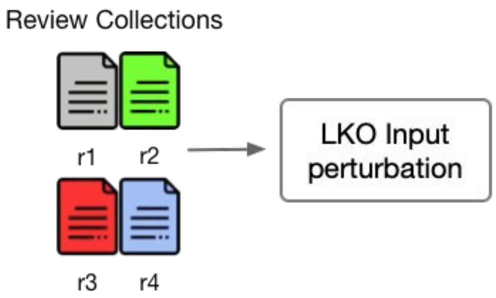
Proposed solution

- **PASS** (Oved and Levi, ACL 2021)
  - **Perturbs** input reviews to generate multiple summaries
  - Ranks them by coherence
- A separate trained coherence model is used for ranking

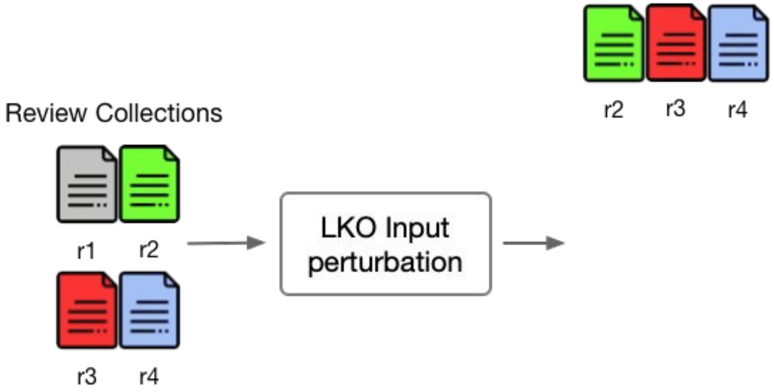
Workflow



Workflow

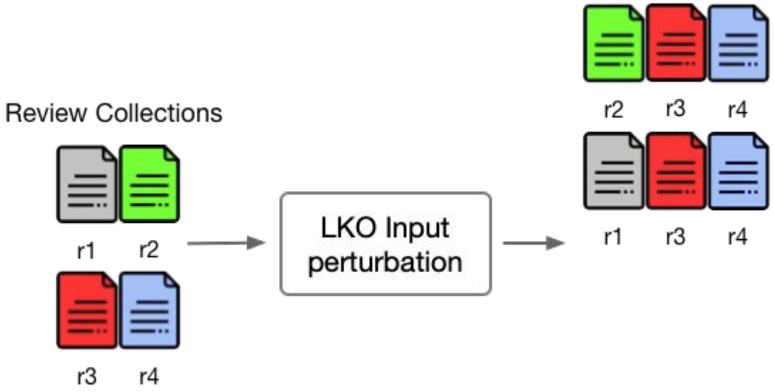


Workflow

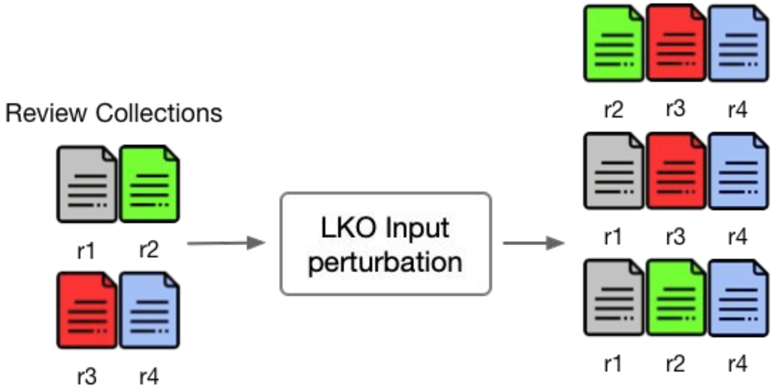


Review Collections
r1
r2
r3
r4
LKO Input perturbation
r2
r3
r4

Workflow

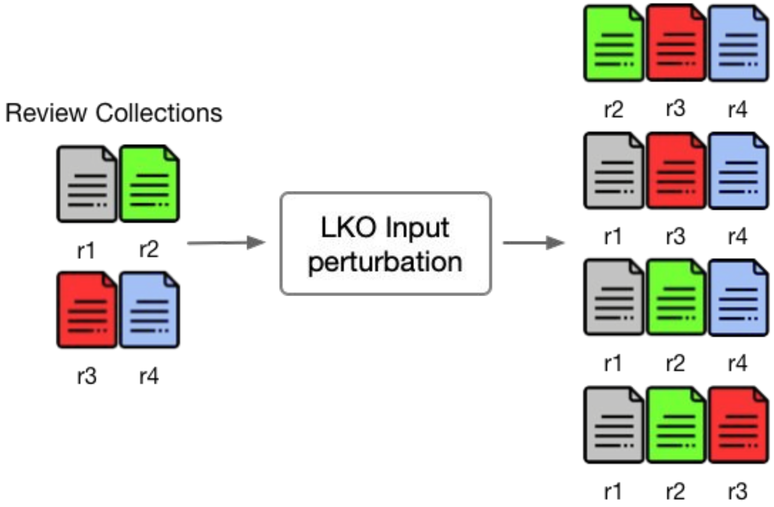

Review Collections
r1
r2
r3
r4
LKO Input perturbation
r2
r3
r4
r1
r3
r4

Workflow

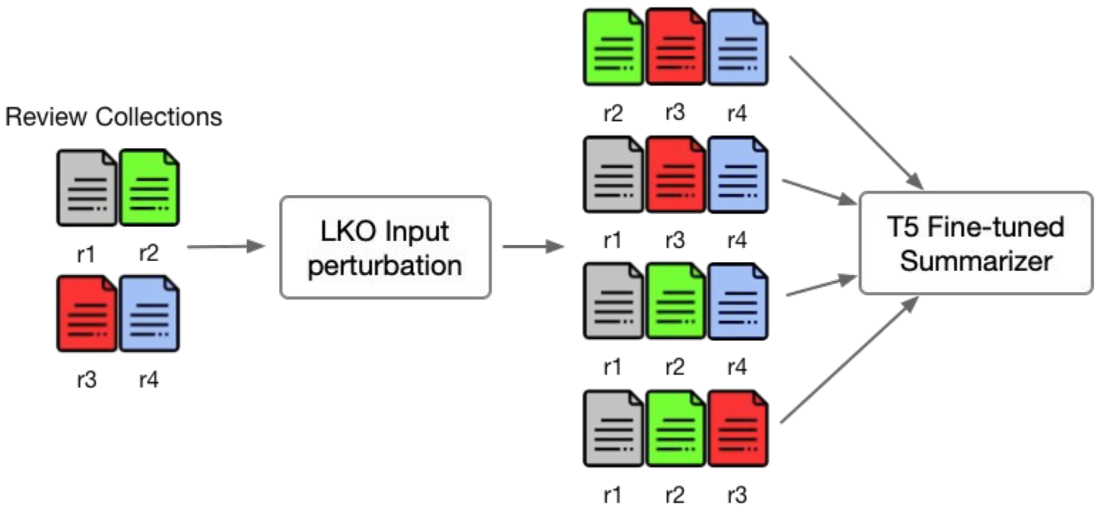


Review Collections
r1
r2
r3
r4
LKO Input perturbation
r2
r3
r4
r1
r3
r4
r1
r2
r4

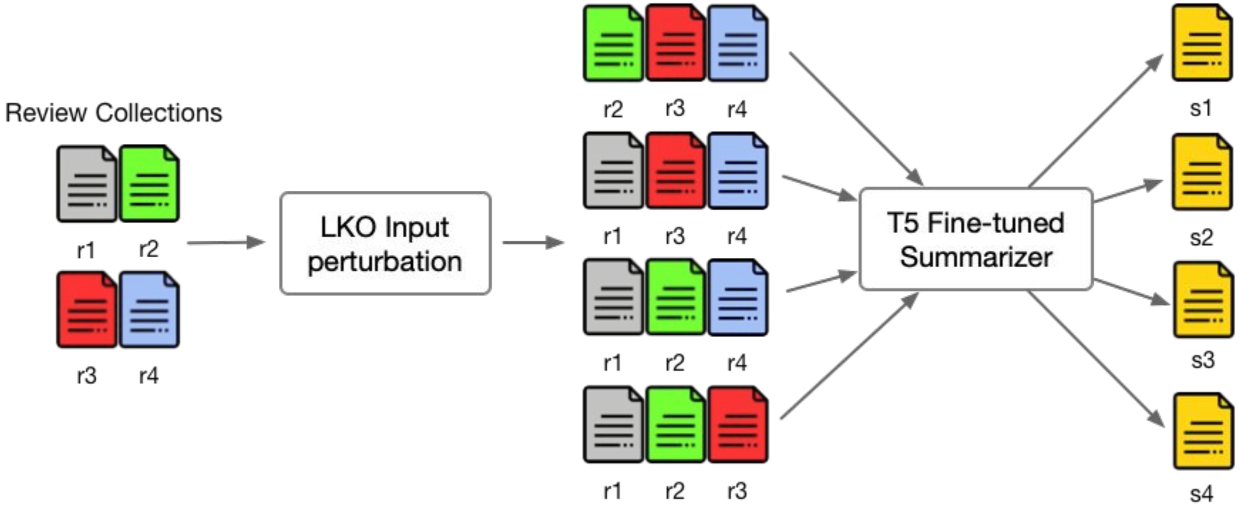
Workflow


Review Collections
r1
r2
r3
r4
LKO Input perturbation
r2
r3
r4
r1
r3
r4
r1
r2
r4
r1
r2
r3

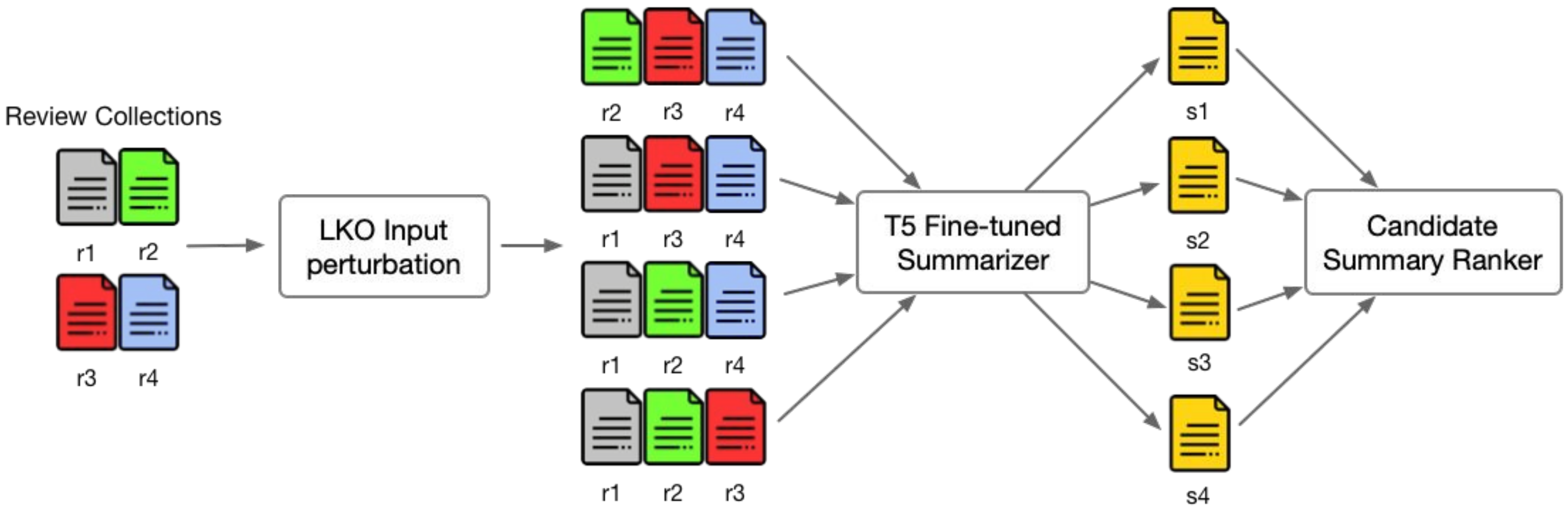
Workflow


Review Collections
r1
r2
r3
r4
LKO Input perturbation
r2
r3
r4
r1
r3
r4
r1
r2
r4
r1
r2
r3
T5 Fine-tuned Summarizer

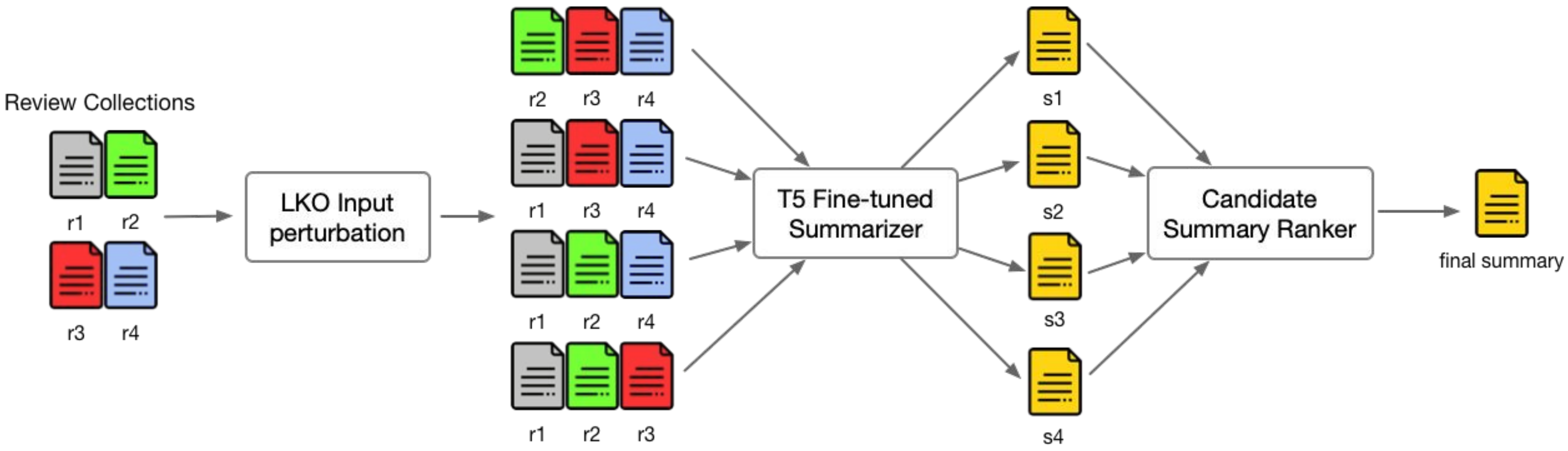
Workflow


Review Collections
r1
r2
r3
r4
LKO Input perturbation
r2
r3
r4
r1
r3
r4
r1
r2
r4
r1
r2
r3
T5 Fine-tuned Summarizer
s1
s2
s3
s4

Workflow


Review Collections
r1
r2
r3
r4
LKO Input perturbation
r2 r3 r4
r1 r3 r4
r1 r2 r4
r1 r2 r3
T5 Fine-tuned Summarizer
s1
s2
s3
s4
Candidate Summary Ranker

Workflow



Review Collections
r1
r2
r3
r4
LKO Input perturbation
r2
r3
r4
r1
r3
r4
r1
r2
r4
r1
r2
r3
T5 Fine-tuned Summarizer
s1
s2
s3
s4
Candidate Summary Ranker
final summary

Candidate Summary Ranker

- **A pairwise** coherence classifier
- Trained on **human annotated news summaries** (Fabbri et al., 2021)
- Count how many times each summary was classified as more coherent
- The final summary is selected based on these counts

Summary examples

FewSum

These running shoes are great! They fit true to size and are very comfortable to run around in. They are light weight and have great support. They run a little on the narrow side, so make sure to order a half size larger than normal.

Summary examples

FewSum

These running shoes are great! They fit true to size and are very comfortable to run around in. They are light weight and have great support. They run a little on the narrow side, so make sure to order a half size larger than normal.

Summary examples

FewSum

These running shoes are great! They fit true to size and are very comfortable to run around in. They are light weight and have great support. They run a little on the narrow side, so make sure to order a half size larger than normal.

PASS

These Reeboks are great for supporting a high arch and are lightweight and comfortable. They come in a variety of colors and sizes, and are ideal for walking or biking. They are also flexible and well made.

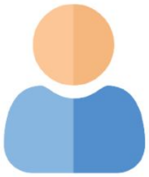
Challenge: Summary Personalization

Why is it challenging?

- No annotated dataset with abstractive summaries
- For example, for **aspect-based summarization**

Aspect-based summarization: illustration

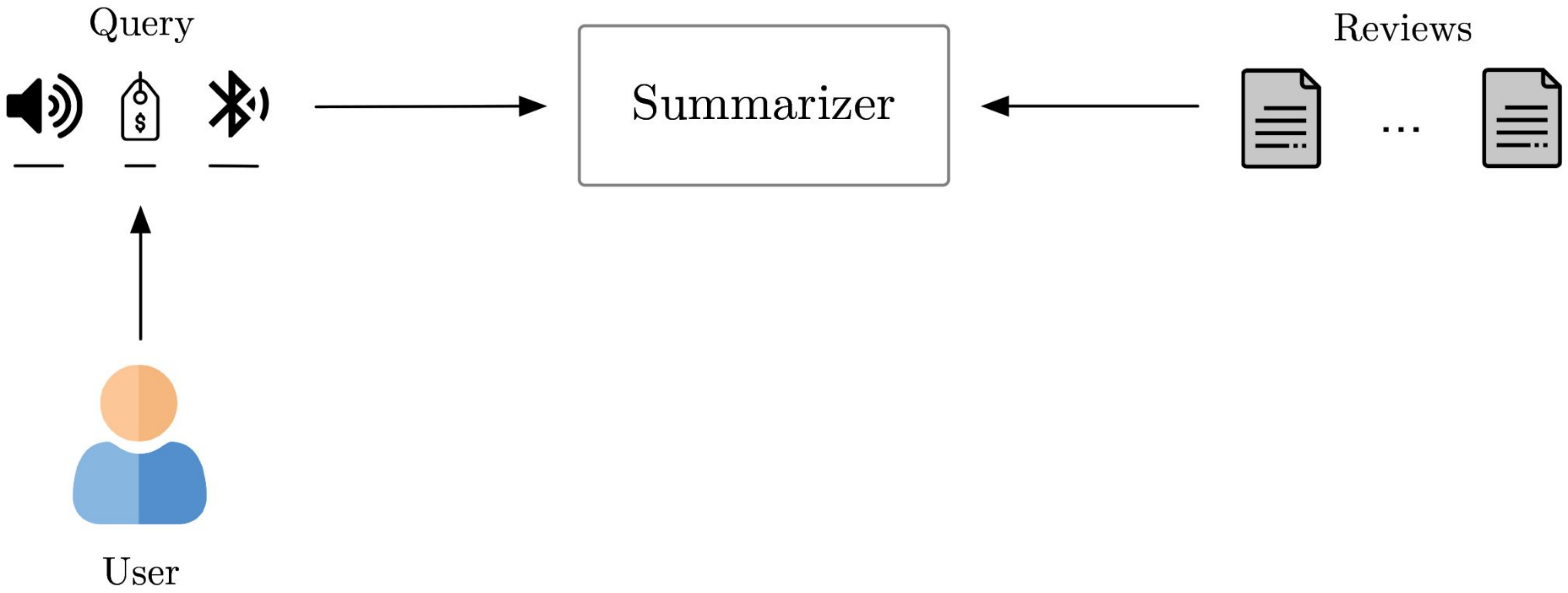
Reviews



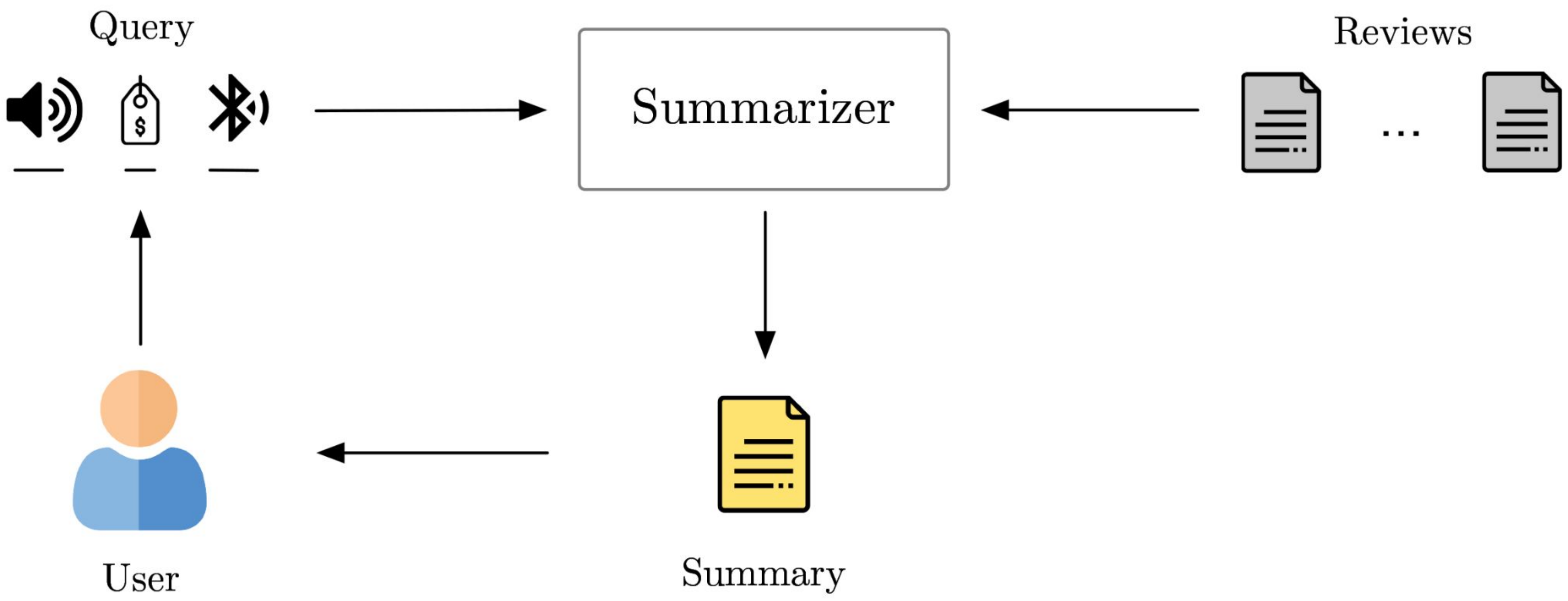
User

Aspect-based summarization: illustration

Aspect-based summarization: illustration



Aspect-based summarization: illustration



Aspect-based summarization

- Aspect queries can be very diverse
- The model needs to **learn to rely on the query**
- Would require **more** annotated samples for **learning** than for **standard summarization**

Proposed solution

- Proposed in AdaQSum (Bražinskas et al., NAACL 2022)
- Similar to AdaSum, it is based on adapters

Proposed solution

- Use an **automatic aspect extractor** to extract fine-grained aspects from target texts
- Use these aspects to construct queries

Proposed solution

The *cover* offers durable *protection* for the MacBook, the retractable *tilt* stands offer *protection* for the *wrists*.

The *keyboard cover* can take some effort to *fit* properly, and *adjustment* to its feel may take time.

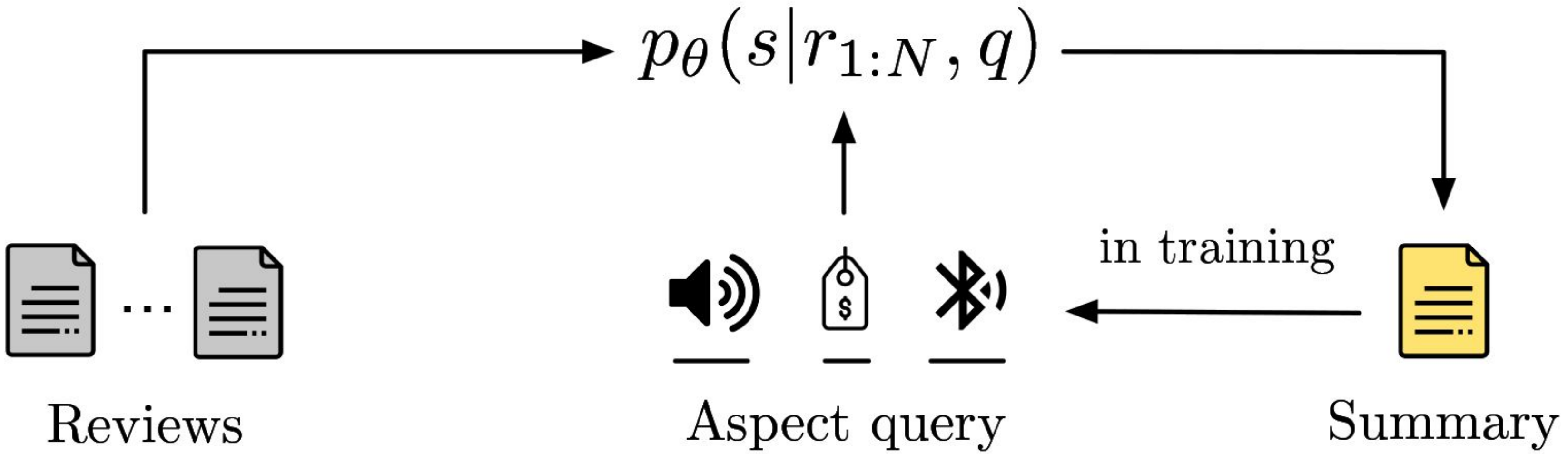
However, free and fast *shipping* make up for this one potential issue.

Table 2: Automatically annotated Amazon summary with fine-grained aspect keywords (*underlined italic*).

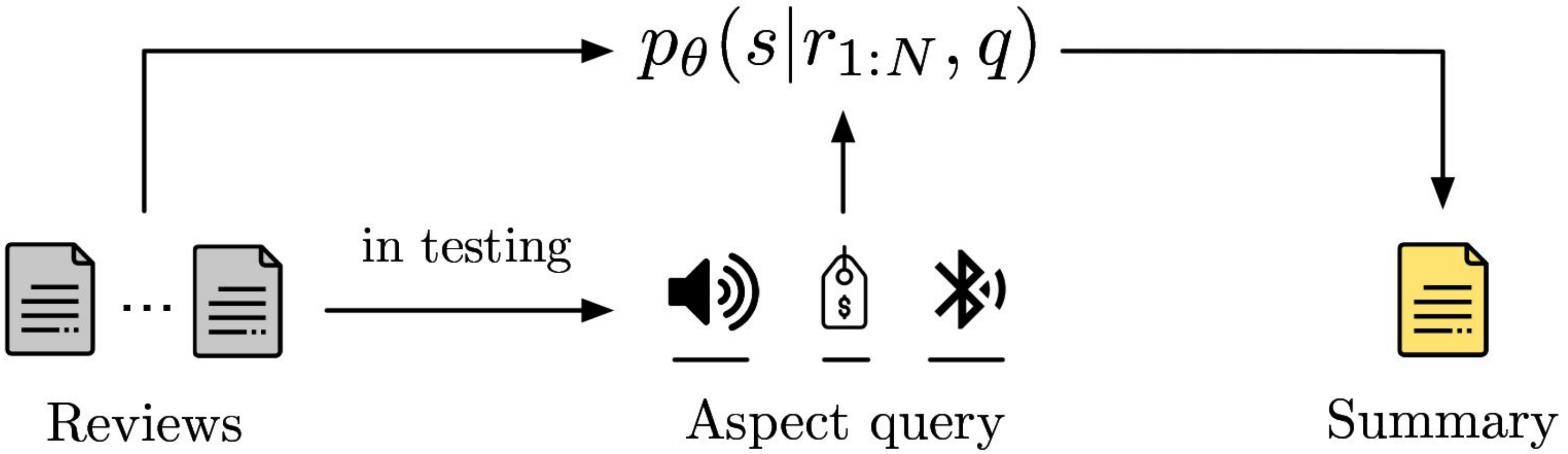
Proposed solution

- Pre-training:
  - Create a synthetic dataset with **pseudo summaries** from customer reviews
  - Create queries for these pseudo summaries
- Fine-tuning:
  - Create queries from **gold summaries**
  - **Fine-tune** on gold samples

Workflow



Workflow



Test time

- Select **top K aspect** keywords from reviews in test time
- Use it to generate summaries

Example summary

Query: fits, weak, color, quality

Example summary

Query: fits, weak, color, quality

Summary: This is a basic tank top that fits well and is comfortable to wear. The color is great and the length is long enough to wear with leggings. The quality of the product is good.

Example summary

Query: fits, weak, color, quality

Summary: This is a basic tank top that fits well and is comfortable to wear. The color is great and the length is long enough to wear with leggings. The quality of the product is good.

Example summary

Query: fits, weak, color, quality

Summary: This is a basic tank top that fits well and is comfortable to wear. The color is great and the length is long enough to wear with leggings. The quality of the product is good.

Example summary

Query: fits, weak, color, quality

Summary: This is a basic tank top that fits well and is comfortable to wear. The color is great and the length is long enough to wear with leggings. The quality of the product is good.

Example summary

Query: fits, weak, color, quality

Summary: This is a basic tank top that fits well and is comfortable to wear. The color is great and the length is long enough to wear with leggings. The quality of the product is good.