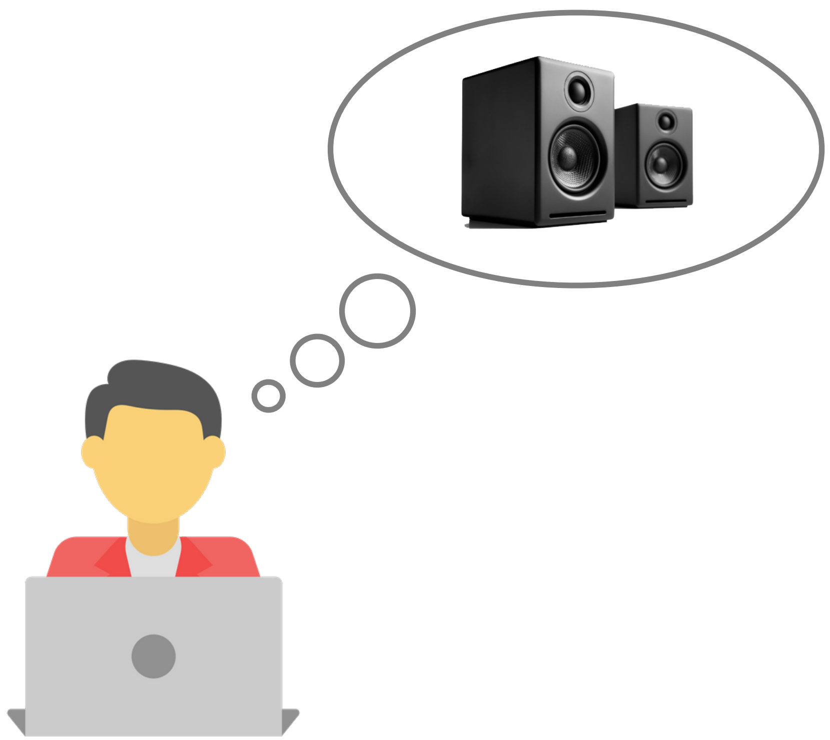
# Unsupervised Opinion Summarization as Copycat-Review Generation

**Arthur Bražinskas**, Mirella Lapata, Ivan Titov
The University of Edinburgh, Scotland

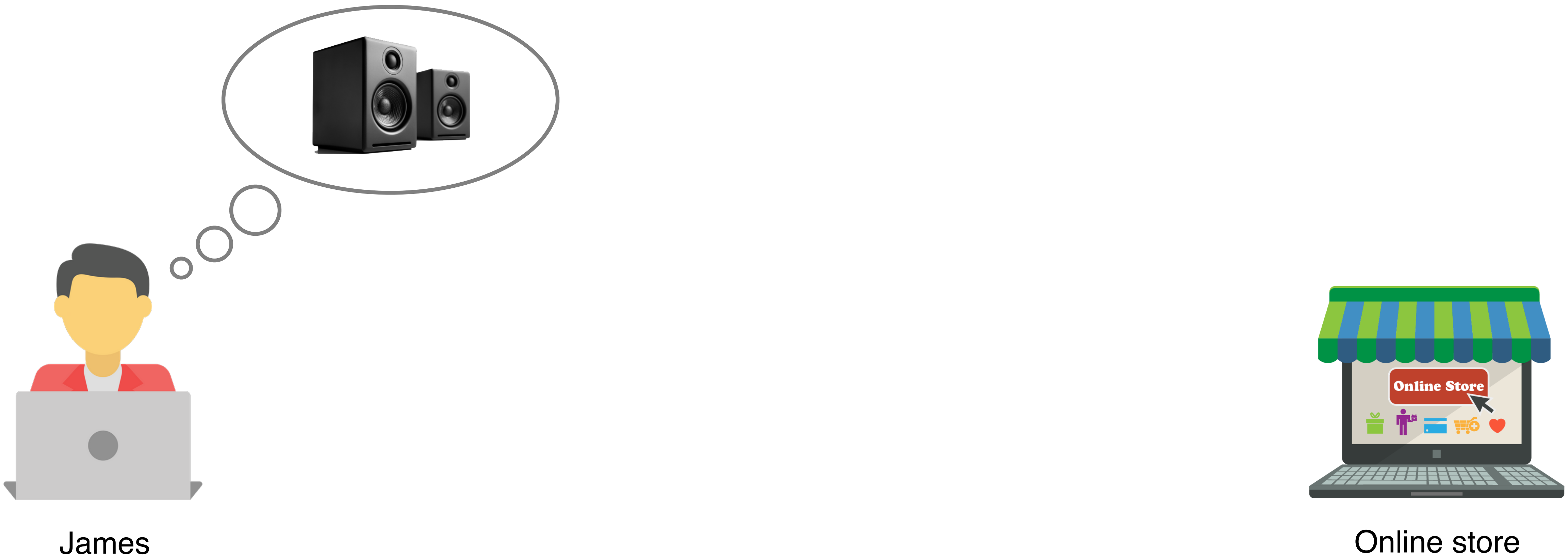
**ACL 2020**

# Opinion Summarization

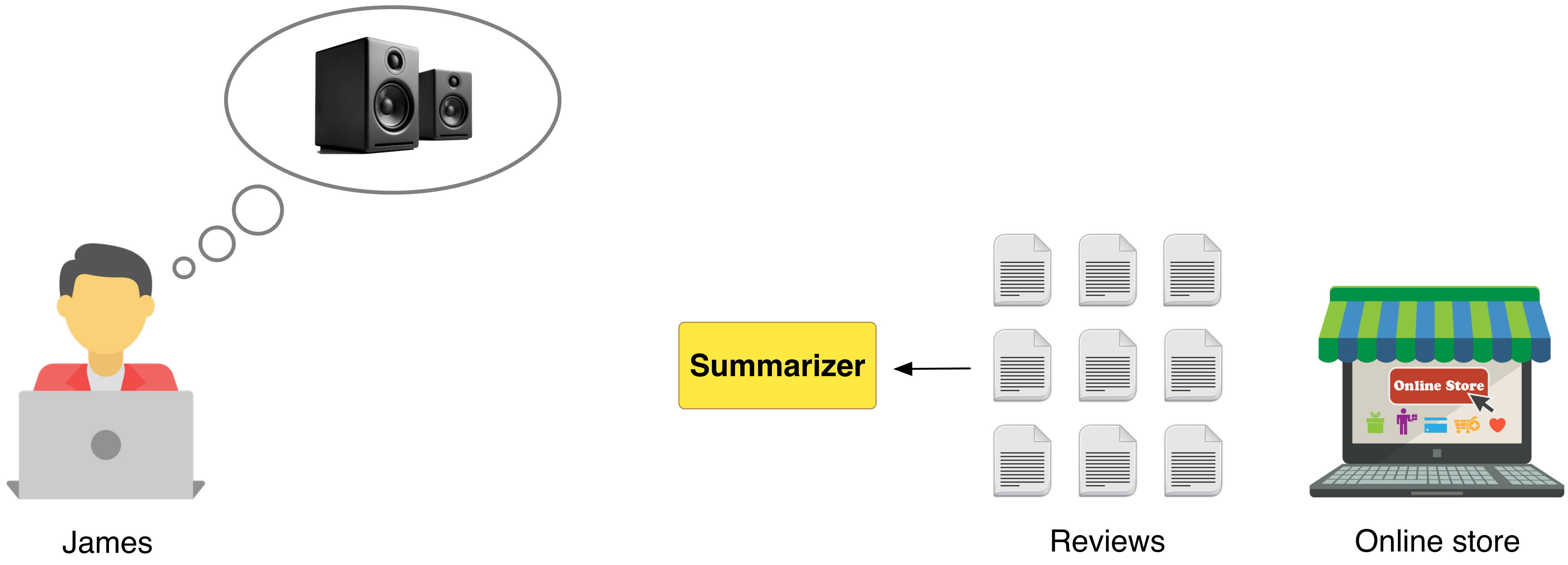
James

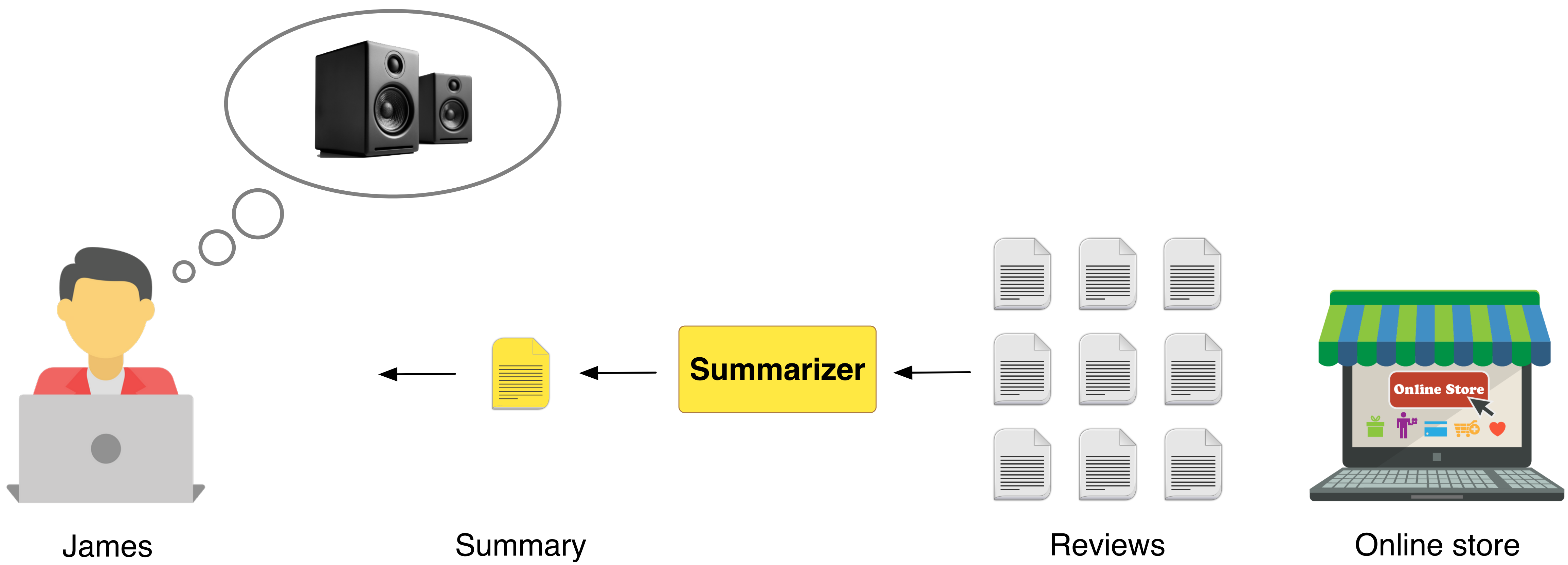


James


James
Online Store
Online store


Online Store
James
Reviews
Online store


Summarizer
Online Store
James
Reviews
Online store


Summarizer
Online Store
James
Summary
Reviews
Online store

# Summarization types

- **Extractive**: select sentences from input documents (e.g., LexRank (Erkan and Radev, 2004))
- **Abstractive**: generate summary text (e.g., MeanSum (Chu and Liu, 2019))
- Most previous works on opinion summarization are **extractive**.
- We introduce an **abstractive summarizer**.

# Example

DAGOSTINO'S

Italian

# Example

The stake was cold, and the bread was sour. The server forgot about our order.

The waitress was very rude. The pasta was too dry, would not recommend it.

# Example

The **stake was cold**, and the **bread was sour**. The **server forgot about our order**.

The **waitress was very rude**. The **pasta was too dry**, would not recommend it.

# Example

The **stake was cold**, and the **bread was sour**. The **server forgot about our order**.

The **waitress was very rude**. The **pasta was too dry**, would not recommend it.

**Extractive summary**: ?

# Example

The **stake was cold**, and the **bread was sour**. The **server forgot about our order**.

**The waitress was very rude.** The **pasta was too dry**, would not recommend it.

**Extractive summary**: The **server forgot about our order**. The **pasta was too dry**, would not recommend it.

# Example

The **stake was cold**, and the **bread was sour**. The **server forgot about our order**.

The **waitress was very rude**. The **pasta was too dry**, would not recommend it.

**Abstractive summary**: Both the **service** and **food** are terrible.

# Advantages of abstractive summarize

- Can use a **richer vocabulary of words**.
- Can **rephrase, condense**, and **abstract**.
- Can deal with **conflicting information**.

# Scarce annotated data

- Datasets with reviews-summary pairs are **very limited**.
- The largest one:**100 pairs with summaries.**
- Large quantities of reviews without summaries (**millions**).

# Our Approach

# Copycat

- **Fully unsupervised**.
- Trained on a **large corpus of reviews** without summaries.

# Conditional LM

- Formulate a **conditional language model (LM)**.
- Predicts a review conditioned on the **other** reviews of a product.

# Conditional LM

$r_2$

| |
|---|
| This ... |

# Conditional LM

$r_2$

This …

This backpack …

…the backpack …

… very sturdy knapsack …

$r_1$

$r_3$

$r_4$

# Conditional LM

$r_2$

This …

This backpack …

$r_1$

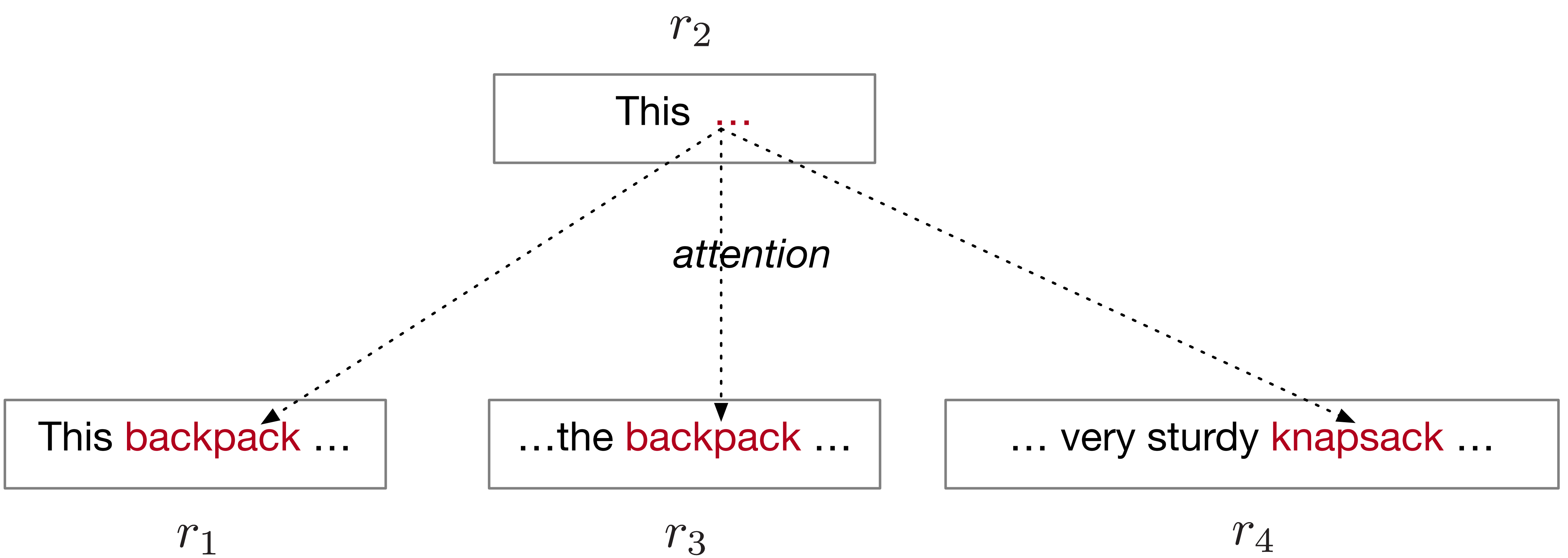
…the backpack …

$r_3$

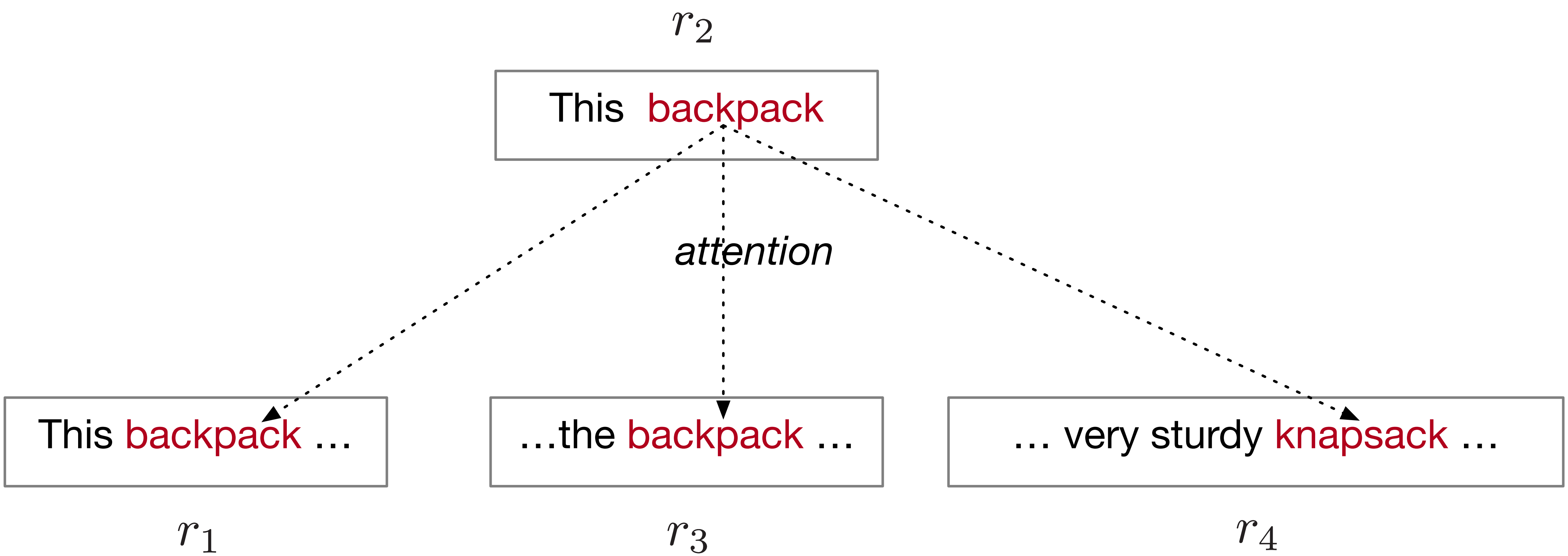
… very sturdy knapsack …

$r_4$

# Conditional LM



# Conditional LM



# Novelty reduction

- Model is trained to predict reviews.
- Summaries are different from reviews in content.
- Summaries do not have **novel content**.
- Control the amount of ‘novelty’ via **latent variables**.

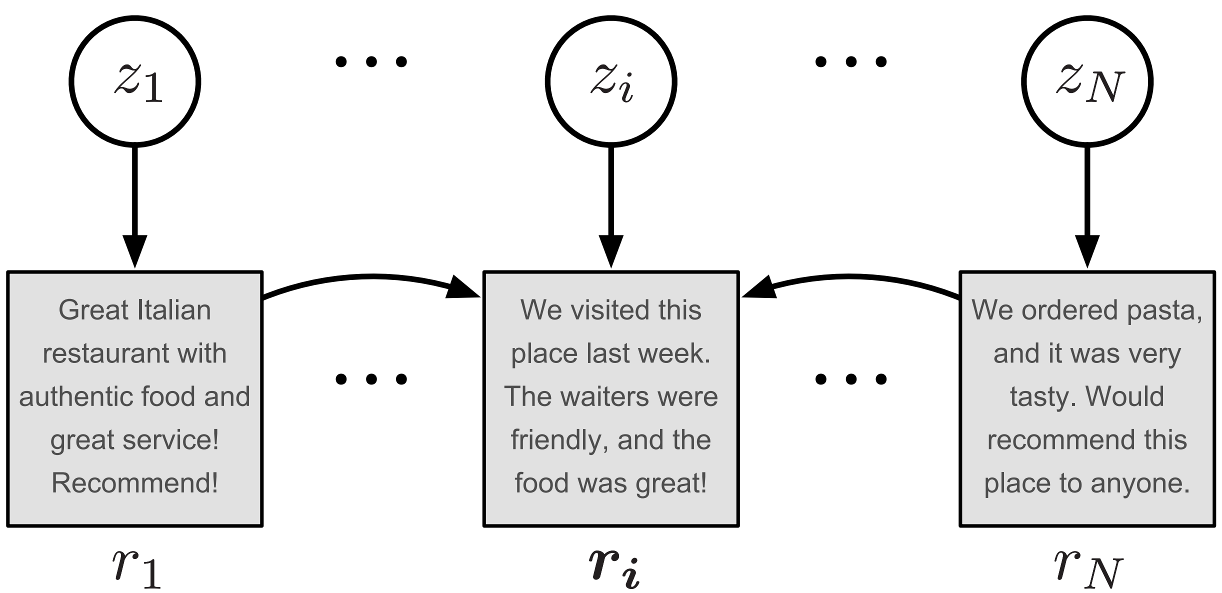
# Latent model


reviews

# Latent model


reviews

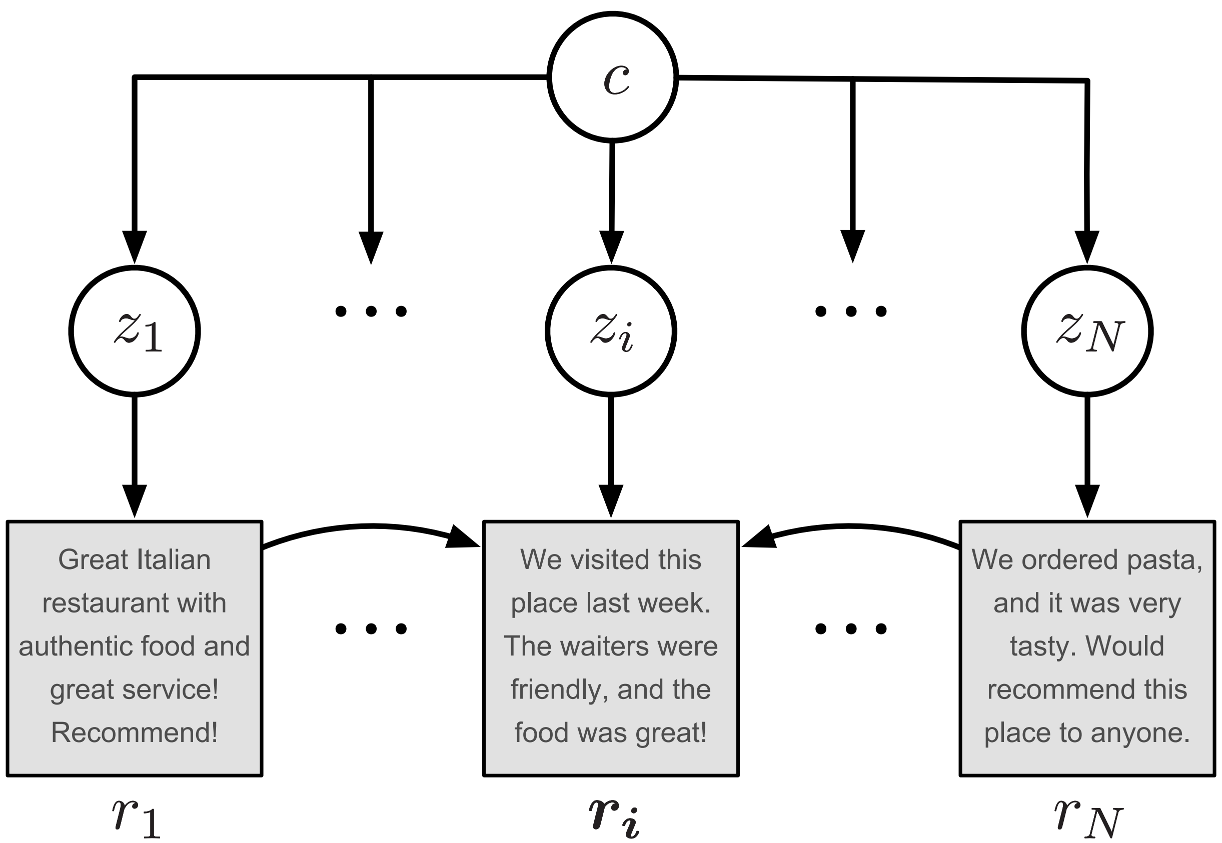
# Latent model




review representations

reviews

# Latent model




product representation

review representations

reviews

# Model training

**Variational Auto-encoders** (Kingma and Welling, 2013) via differentiable sampling.

# Summary generation

- Use **mean values** of the latent variables to **limit novelty**.
- Show that they correspond to **summarizing reviews**.

# Summary generation

1. Infer **the mean** representation of the product:

$$c^* = \mathbb{E}_{c \sim q_\phi(c|r_{1:N})}[c]$$

# Summary generation

1. Infer **the mean** representation of the product:

$$c^* = \mathbb{E}_{c \sim q_\phi(c|r_{1:N})}[c]$$

2. Infer **the mean** representation of the review:

$$z^* = \mathbb{E}_{z \sim p(z|c^*)}[z]$$

# Summary generation

1. Infer **the mean** representation of the product:

$$c^* = \mathbb{E}_{c \sim q_\phi(c|r_{1:N})}[c]$$

2. Infer **the mean** representation of the review:

$$z^* = \mathbb{E}_{z \sim p(z|c^*)}[z]$$

3. Generate **the summarizing review**:

$$r^* = \arg\max_r p_\theta(r|z^*, r_{1:N})$$

# Example Summary

| | |
|---|---|
| **Summary** | This restaurant is a hidden gem in Toronto. The food is delicious, and the service is impeccable. Highly recommend for anyone who likes French bistro. |
| **Reviews** | We got the steak frites and the chicken frites both of which were very good ... Great service ... \|\| I really love this place ... Côte de Boeuf ... A Jewel in the big city ... \|\| French jewel of Spadina and Adelaide , Jules ... They are super accommodating ... moules and frites are delicious ... \|\| Food came with tons of greens and fries along with my main course , thumbs uppp ... \|\| Chef has a very cool and fun attitude ... \|\| Great little French Bistro spot ... Go if you want French bistro food classics ... \|\| Great place ... the steak frites and it was amazing ... Best Steak Frites ... in Downtown Toronto ... \|\| Favourite french spot in the city ... crème brule for dessert |

| | |
|---|---|
| **Summary** | This restaurant is a hidden gem in Toronto. The food is delicious, and the service is impeccable. Highly recommend for anyone who likes French bistro. |
| **Reviews** | We got the steak frites and the chicken frites both of which were very good ... Great service ... \|\| I really love this place ... Côte de Boeuf ... A Jewel in the big city ... \|\| French jewel of Spadina and Adelaide , Jules ... They are super accommodating ... moules and frites are delicious ... \|\| Food came with tons of greens and fries along with my main course , thumbs uppp ... \|\| Chef has a very cool and fun attitude ... \|\| Great little French Bistro spot ... Go if you want French bistro food classics ... \|\| Great place ... the steak frites and it was amazing ... Best Steak Frites ... in Downtown Toronto ... \|\| Favourite french spot in the city ... crème brule for dessert |

| | |
|---|---|
| **Summary** | This restaurant is a hidden gem in Toronto. The food is delicious, and the service is impeccable. Highly recommend for anyone who likes French bistro. |
| **Reviews** | We got the steak frites and the chicken frites both of which were very good ... Great service ... \|\| I really love this place ... Côte de Boeuf ... A Jewel in the big city ... \|\| French jewel of Spadina and Adelaide , Jules ... They are super accommodating ... moules and frites are delicious ... \|\| Food came with tons of greens and fries along with my main course , thumbs uppp ... \|\| Chef has a very cool and fun attitude ... \|\| Great little French Bistro spot ... Go if you want French bistro food classics ... \|\| Great place ... the steak frites and it was amazing ... Best Steak Frites ... in Downtown Toronto ... \|\| Favourite french spot in the city ... crème brule for dessert |

| | |
|---|---|
| **Summary** | This restaurant is a hidden gem in Toronto. The food is delicious, and the service is impeccable. Highly recommend for anyone who likes French bistro. |
| **Reviews** | We got the steak frites and the chicken frites both of which were very good ... Great service ... \|\| I really love this place ... Côte de Boeuf ... A Jewel in the big city ... \|\| French jewel of Spadina and Adelaide , Jules ... They are super accommodating ... moules and frites are delicious ... \|\| Food came with tons of greens and fries along with my main course , thumbs uppp ... \|\| Chef has a very cool and fun attitude ... \|\| Great little French Bistro spot ... Go if you want French bistro food classics ... \|\| Great place ... the steak frites and it was amazing ... Best Steak Frites ... in Downtown Toronto ... \|\| Favourite french spot in the city ... crème brule for dessert |

# Experiments

# Data

- Unannotated review datasets:
  - Yelp: 1M
  - Amazon: 4.5M (He and McAuley, 2016)
- Tested on human-written summaries.
- Also, annotated 180 summaries for Amazon products.

# Results

- Show truncated results on Yelp.
- Amazon results are very similar (see the paper).

# ROUGE

- ROUGE is the **central metric** for automatic evaluation.
- Based on **n-gram overlap** between a generated and true summary.

# ROUGE

| | R1 | R2 | RL |
|---|---|---|---|
| | | | |
| | | | |

# ROUGE

| | R1 | R2 | RL |
|---|---|---|---|
| Clustroid | 0.2628 | 0.0348 | 0.1536 |
| Lead | 0.2634 | 0.0372 | 0.1386 |
| Random | 0.2304 | 0.0244 | 0.1344 |

# ROUGE

| | R1 | R2 | RL |
|---|---|---|---|
| LexRank | 0.2501 | 0.0362 | 0.1467 |
| Clustroid | 0.2628 | 0.0348 | 0.1536 |
| Lead | 0.2634 | 0.0372 | 0.1386 |
| Random | 0.2304 | 0.0244 | 0.1344 |

# ROUGE

| | R1 | R2 | RL |
|---|---|---|---|
| MeanSum | 0.2846 | 0.0366 | 0.1557 |
| LexRank | 0.2501 | 0.0362 | 0.1467 |
| Clustroid | 0.2628 | 0.0348 | 0.1536 |
| Lead | 0.2634 | 0.0372 | 0.1386 |
| Random | 0.2304 | 0.0244 | 0.1344 |

# ROUGE

| | R1 | R2 | RL |
|---|---|---|---|
| Copycat | **0.2947** | **0.0526** | **0.1809** |
| MeanSum | 0.2846 | 0.0366 | 0.1557 |
| LexRank | 0.2501 | 0.0362 | 0.1467 |
| Clustroid | 0.2628 | 0.0348 | 0.1536 |
| Lead | 0.2634 | 0.0372 | 0.1386 |
| Random | 0.2304 | 0.0244 | 0.1344 |

# Content support

- Abstractive systems can produce content that is **not aligned** with input reviews.
- E.g., ‘iPhone’ instead of ‘iPad’.
- False content can lead to user aversion.

# Content support

- Split **Copycat's** and **MeanSum's** summaries by sentences.

- Hired AMT workers to judge how well summary sentence content is supported.

# Content support

| | Full (%) | Partial (%) | No (%) |
|---|---|---|---|
| Copycat | **44.50** | **32.48** | **23.01** |
| MeanSum | 28.41 | 30.66 | 40.92 |

# Conclusion

- **Unsupervised** abstractive summarization model.
- **Control of novelty** via latent variables.
- Tackling of **under-explored** abstractive opinion summarization.
- **Strong results** in evaluation.

# More in the paper

- Detailed ablation (e.g., over latent variables).
  - Show that the latent variables are essential.
- Additional human evaluations (Best-Worst scaling).
- Analysis of the summary difference when sampling and mean values of latent variables are used.

# Limitations

- Summation is limited to 8 reviews.
- Summaries can have the writing style of a review.
- Consensus summaries do not contrast opinions.

<END>