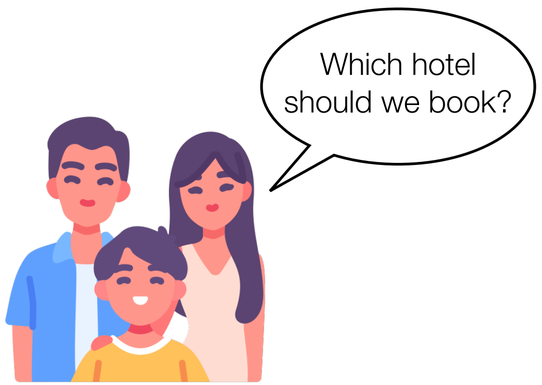
UNC NLP

Unsupervised Extractive Opinion Summarization Using Sparse Coding

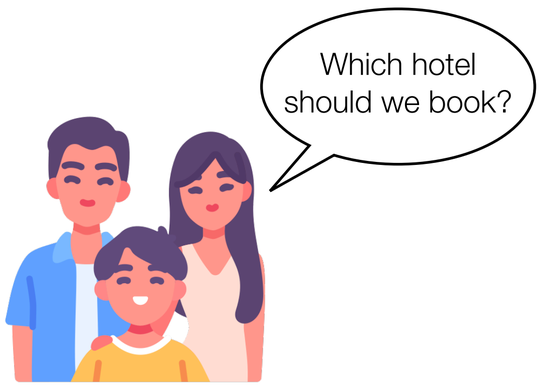
Somnath Basu Roy Chowdhury, *Chao Zhao and Snigdha Chaturvedi*

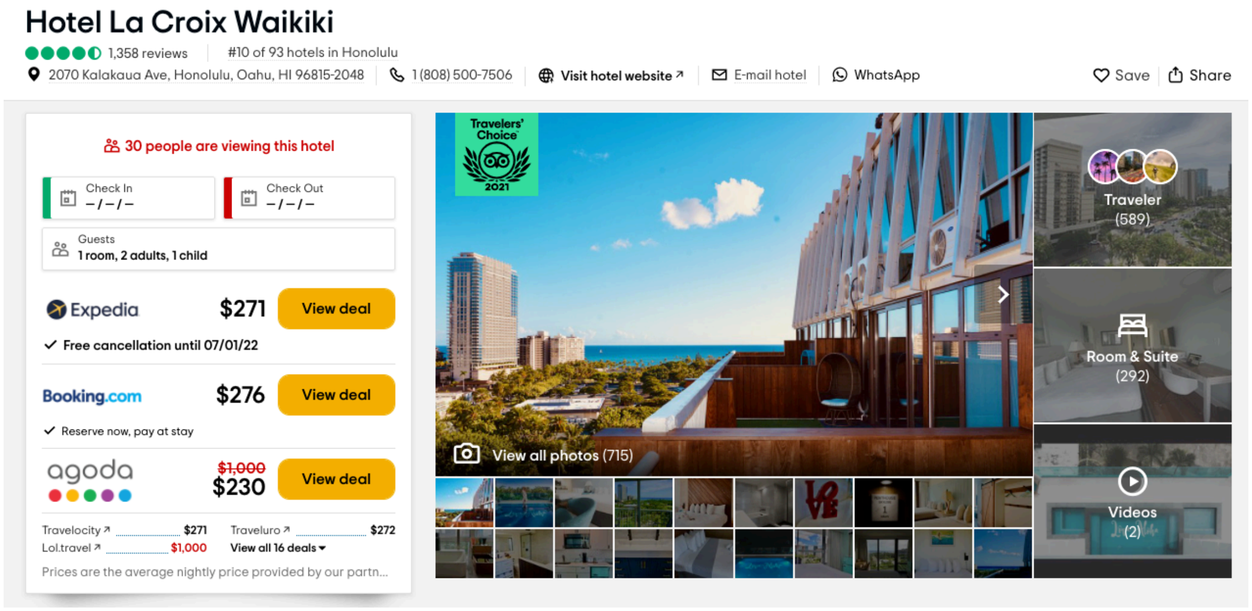
ACL 2022

Opinion Summarization

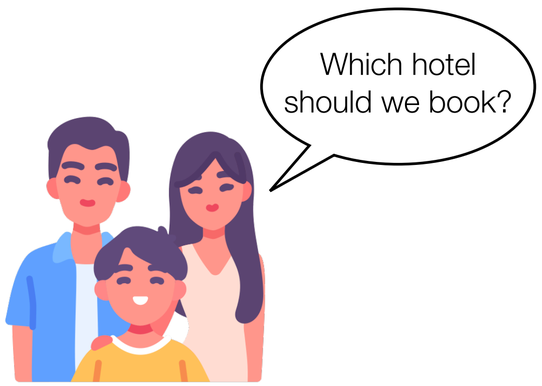


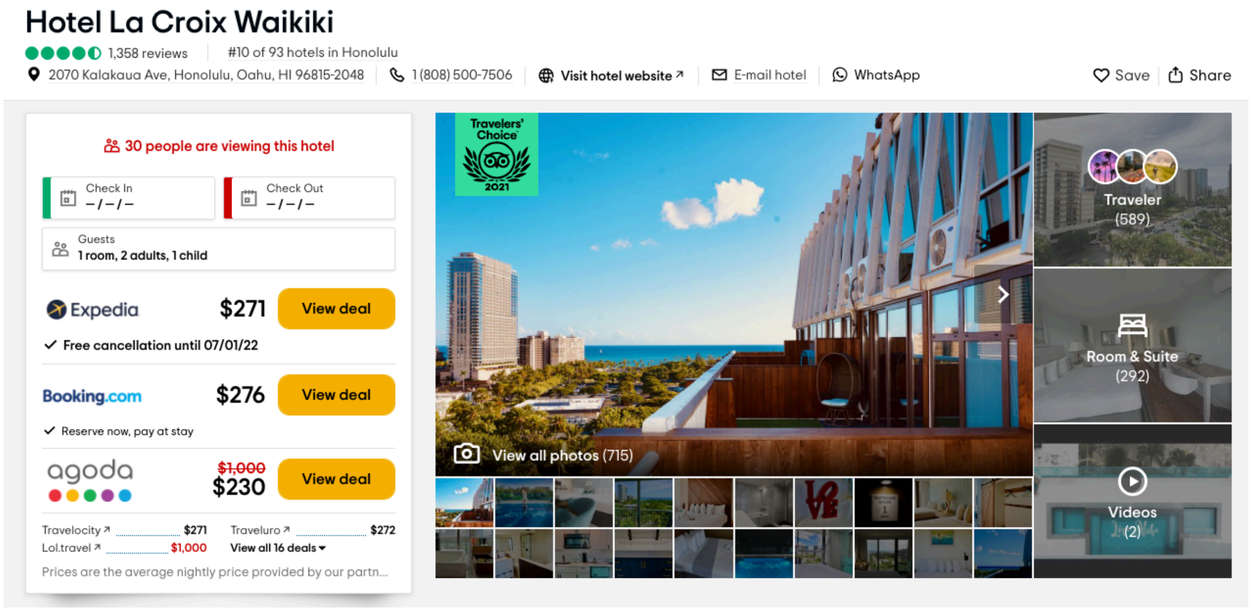
Opinion Summarization



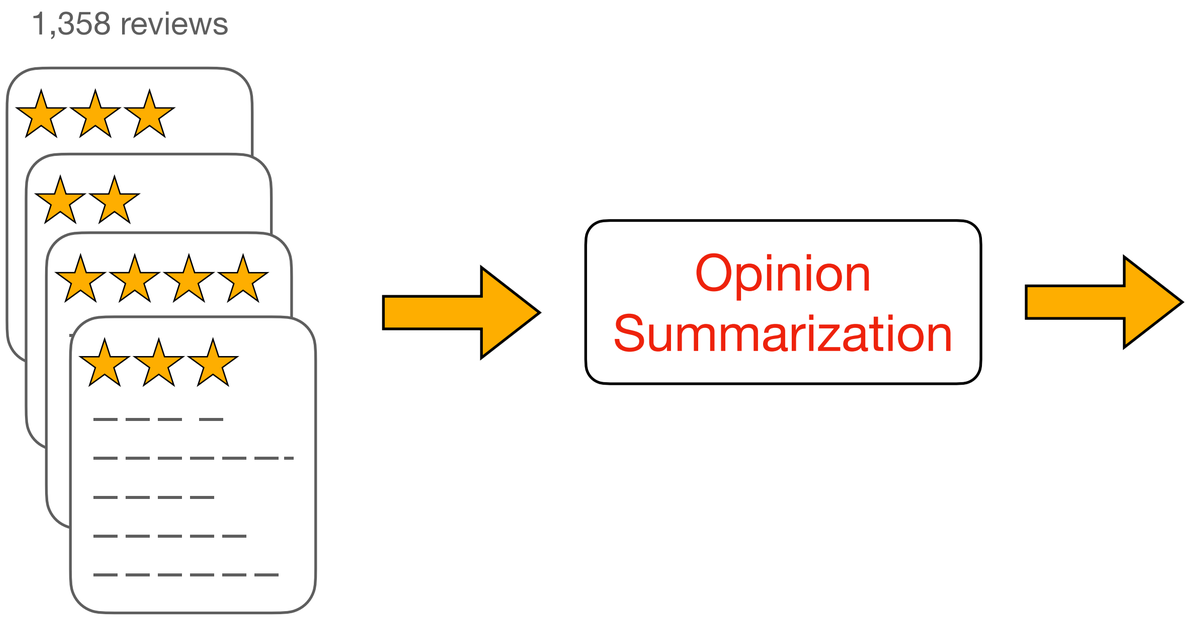


Opinion Summarization

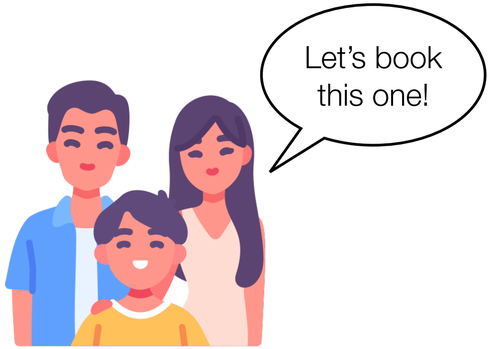




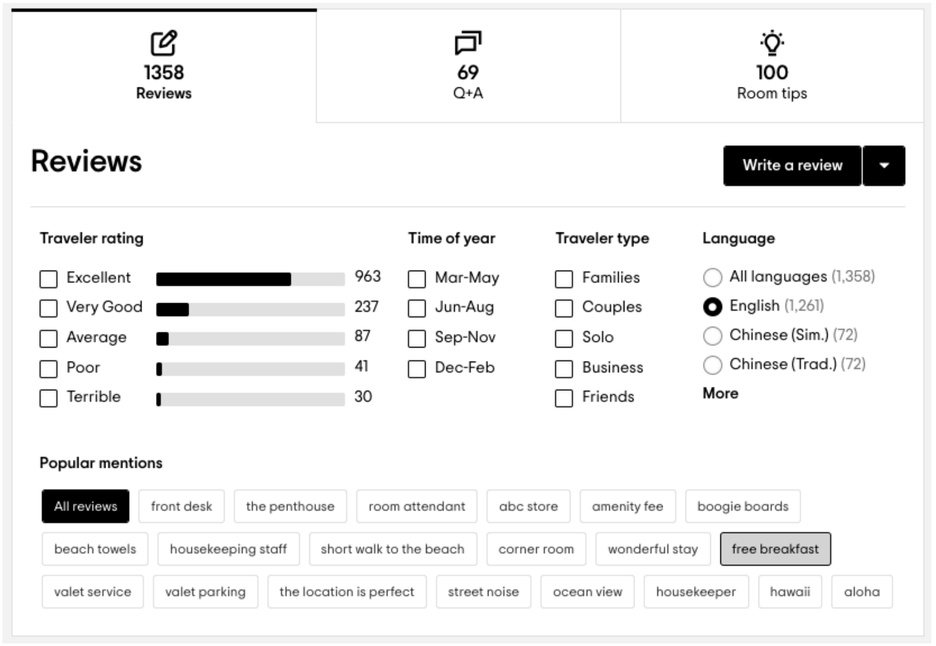
Opinion Summarization



Opinion Summarization



Opinion Summarization



Motivation

- Unsupervised opinion systems are desirable due to the scarcity of labeled data

Motivation

- Unsupervised opinion systems are desirable due to the scarcity of labeled data
- It is crucial to understand the underlying semantics in an opinion

Motivation

- Unsupervised opinion systems are desirable due to the scarcity of labeled data
- It is crucial to understand the underlying semantics in an opinion
- It can be captured as a distribution over latent semantic units

Motivation

- Unsupervised opinion systems are desirable due to the scarcity of labeled data
- It is crucial to understand the underlying semantics in an opinion
- It can be captured as a distribution over latent semantic units
- Opinions aligning with popular semantic distribution are selected

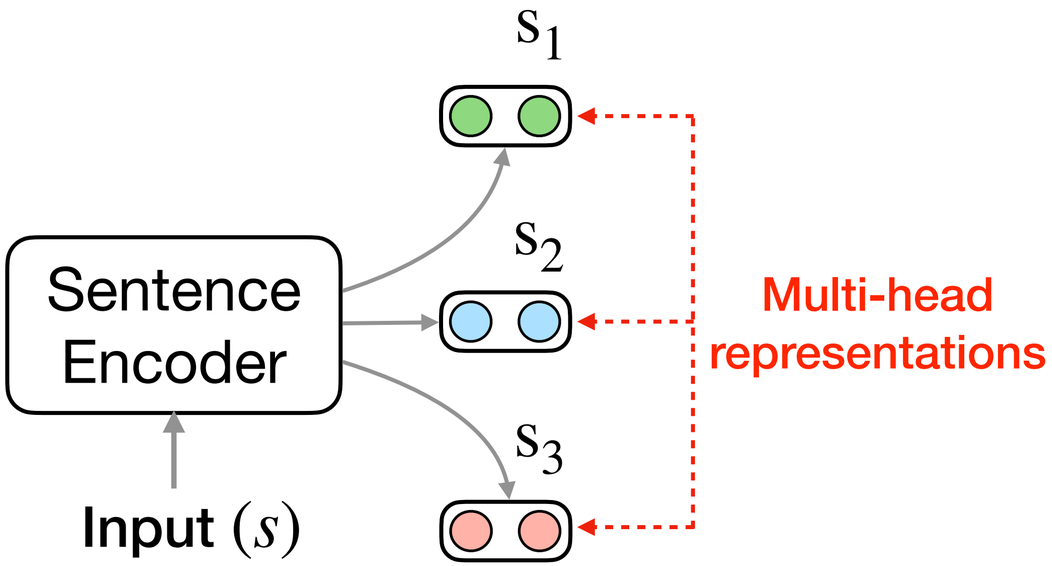
Semantic Autoencoder

Performs summarization in the following stages:

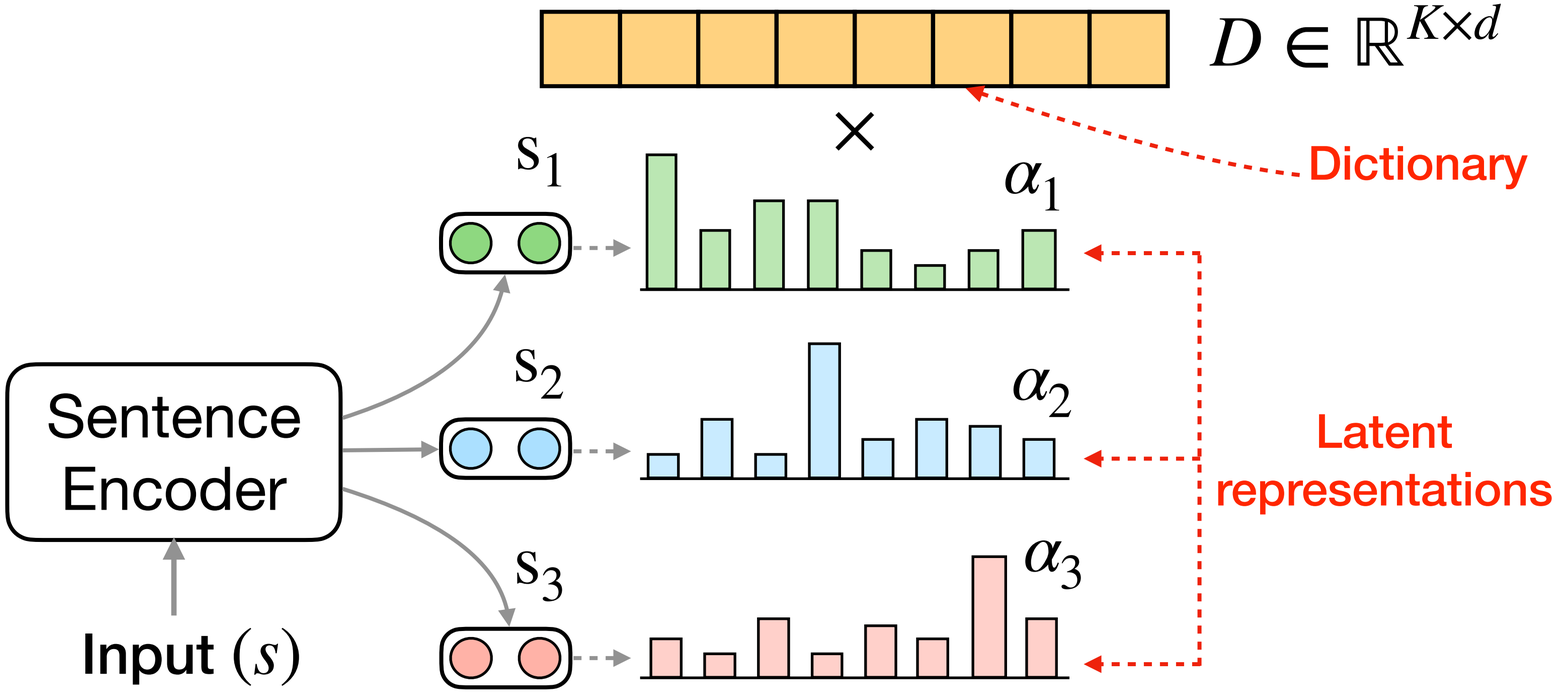
- Text Representation Learning
- Summarization based on saliency scores
  - General Summarization — **relevance**, **redundancy** and **aspect-awareness**
  - Aspect Summarization — **relevance** and **informativeness**

Text Representation Learning

Text Representation Learning

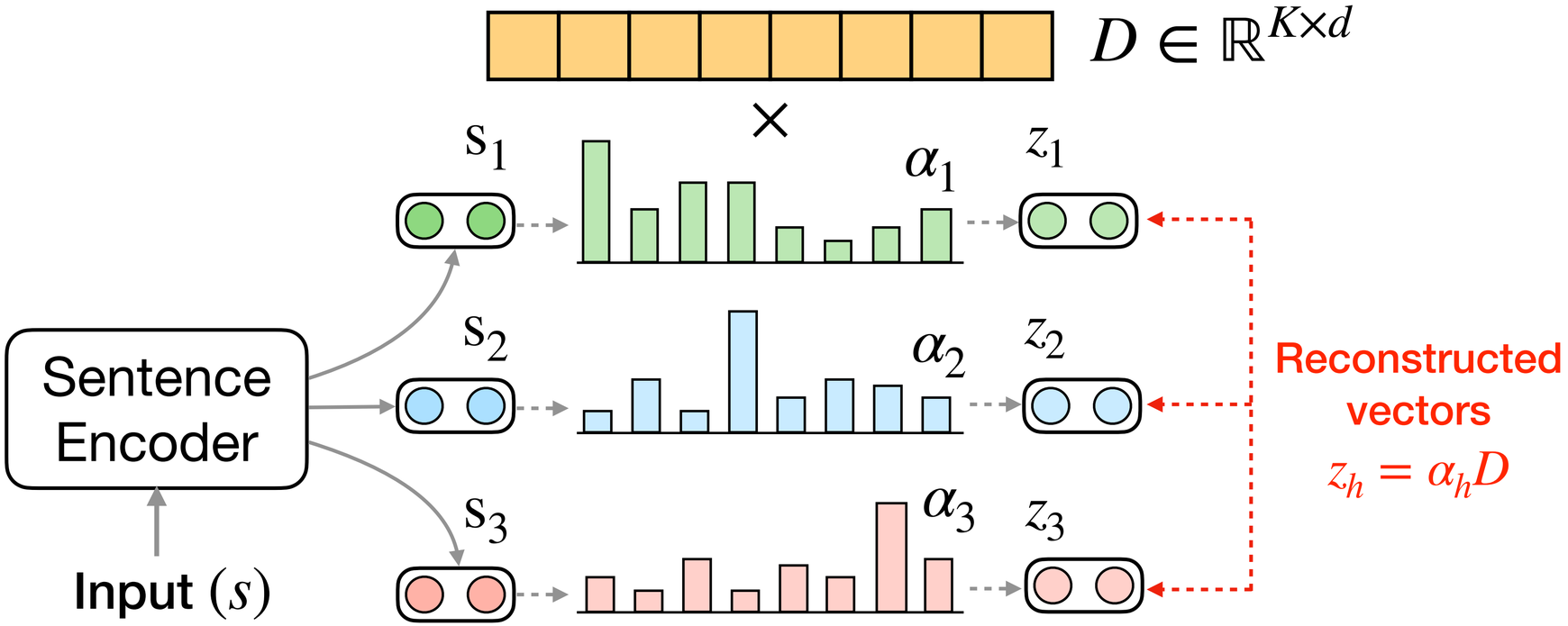


Text Representation Learning

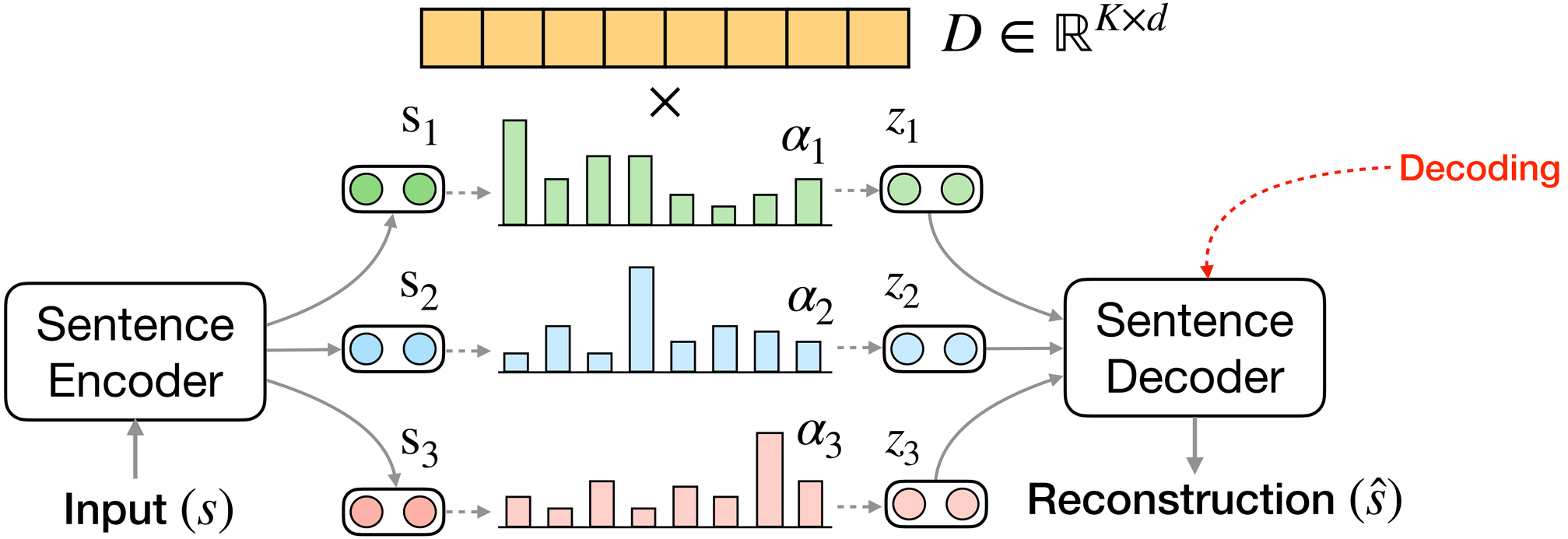


$$\alpha_h = \text{softmax}(s_h D^T)$$

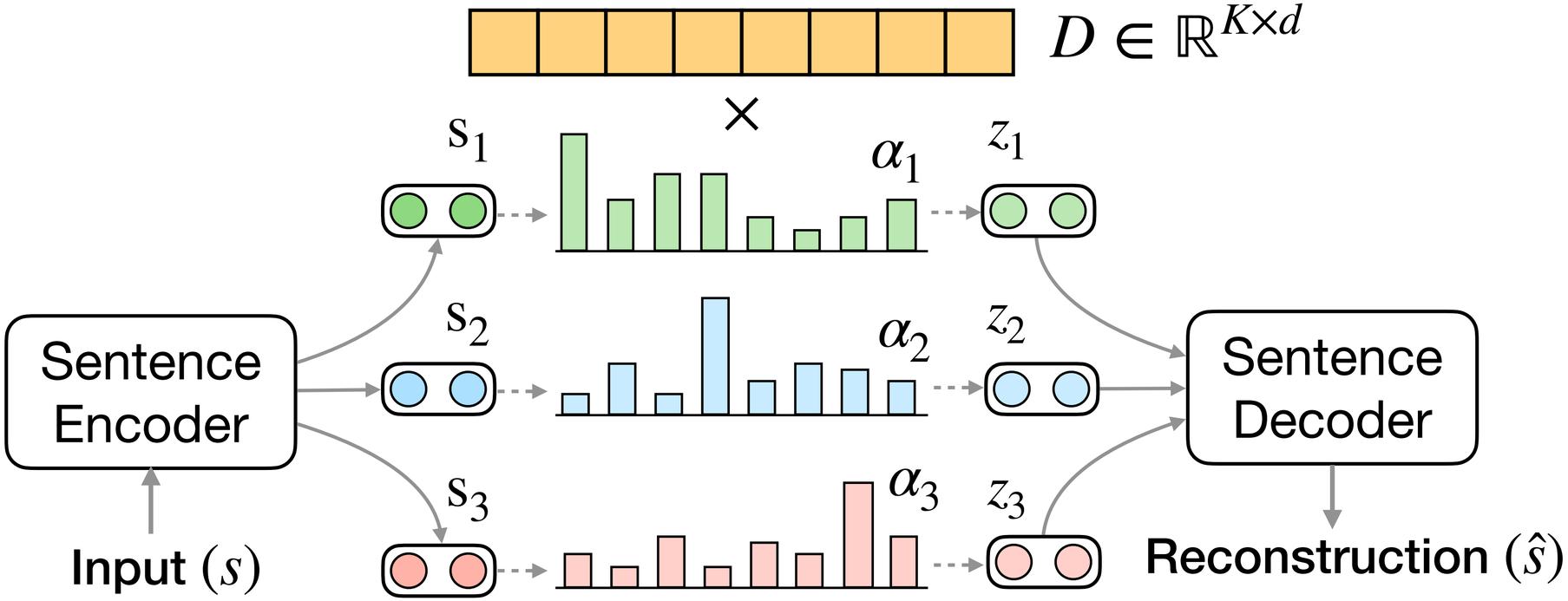
Text Representation Learning



Text Representation Learning



Training



Sparsity promoting losses

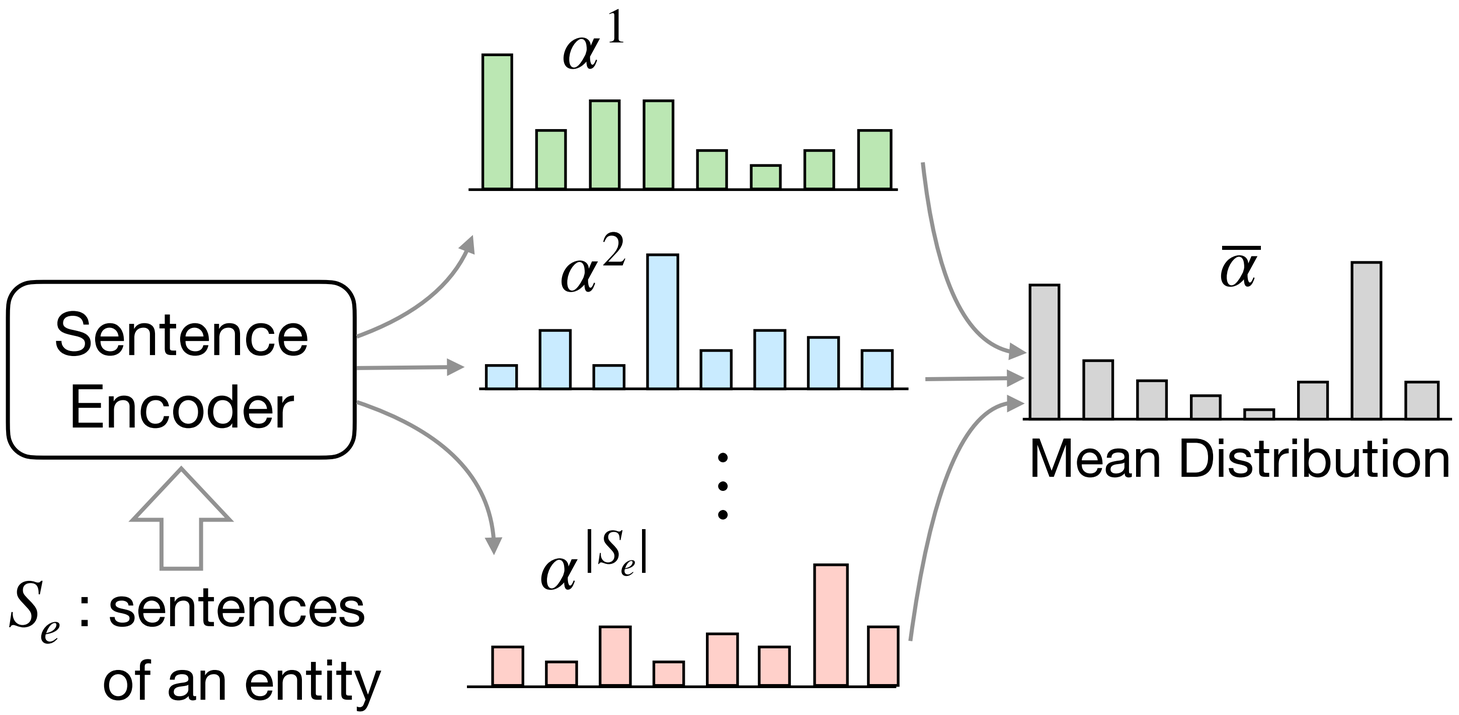
$$\lambda_1 \sum_h |\alpha_h| + \lambda_2 \sum_h H(\alpha_h)$$

Decoder loss

$$\mathscr{L}_{\text{CE}}(s, \hat{s})$$

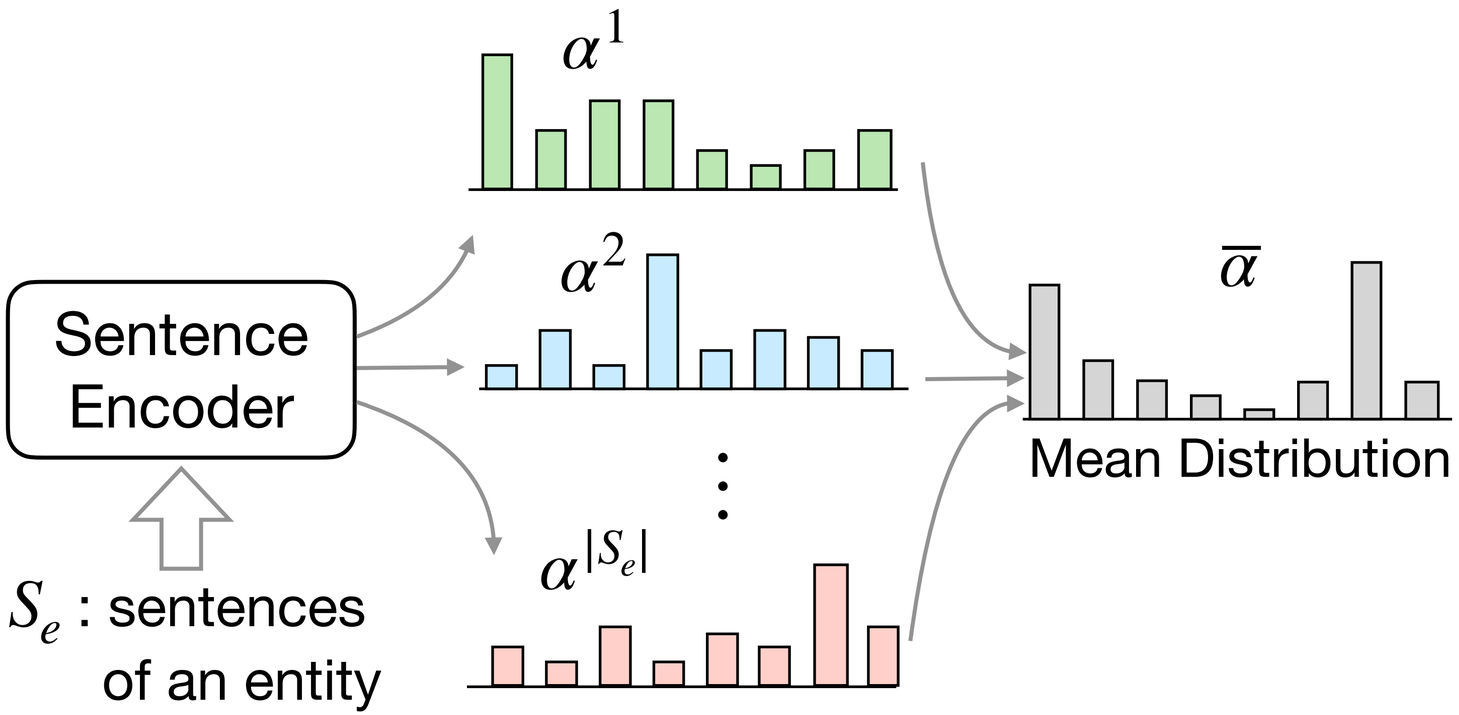
Sentence Selection

Summarization using α_h



Saliency score $\mathscr{R}(\alpha^s)$:

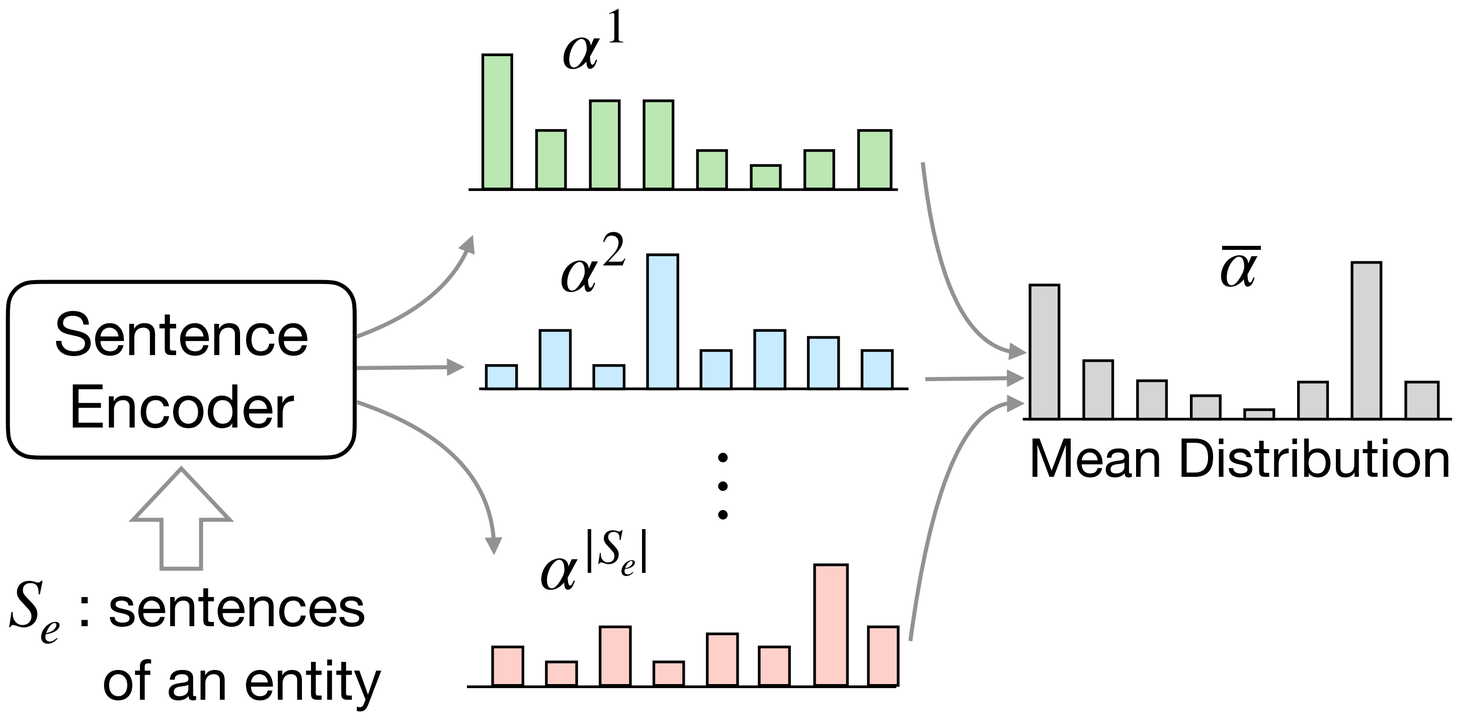
Summarization using α_h



Saliency score $\mathcal{R}(\alpha^s)$:

- **Relevance:** $\Delta(\bar{\alpha}, \alpha^s)$

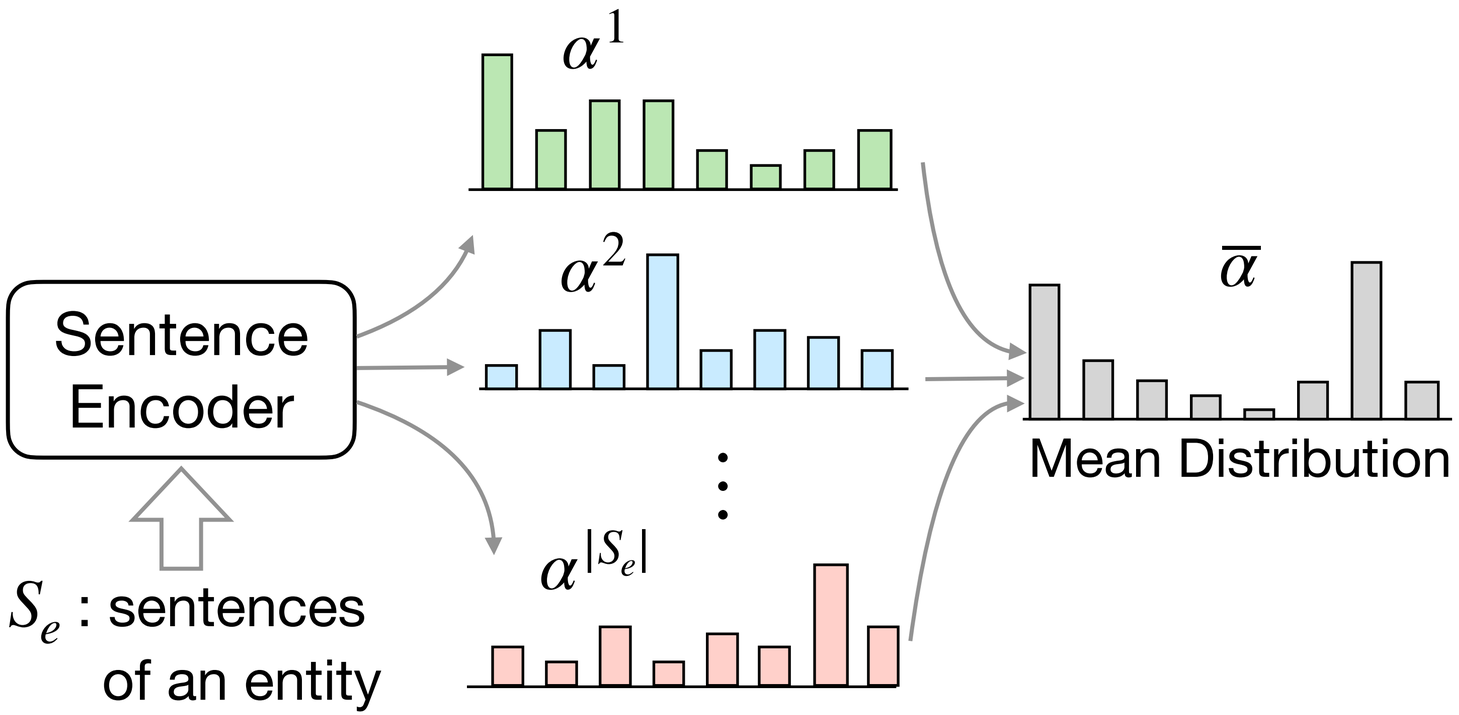
Summarization using α_h



Saliency score $\mathcal{R}(\alpha^s)$:

- **Relevance:** $\Delta(\bar{\alpha}, \alpha^s)$
- **Redundancy:** $-\gamma \max_{s' \in \hat{O}_e} \Delta(\alpha^{s'}, \alpha^s)$

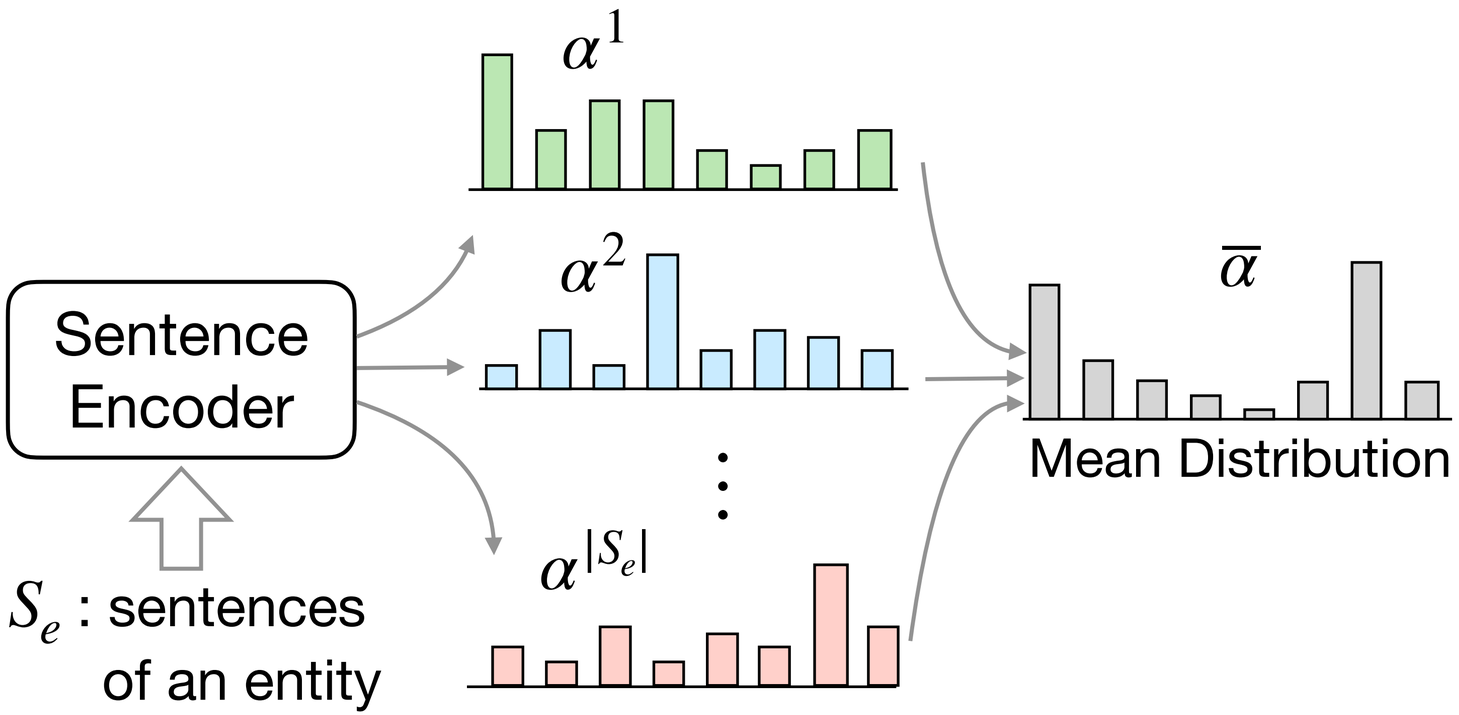
Summarization using α_h



Saliency score $\mathcal{R}(\alpha^s)$:

- **Relevance:** $\Delta(\bar{\alpha}, \alpha^s)$
- **Redundancy:** $-\gamma \max_{s' \in \hat{O}_e} \Delta(\alpha^{s'}, \alpha^s)$
- **Aspect-awareness:** Iterate over aspects and select salient sentences

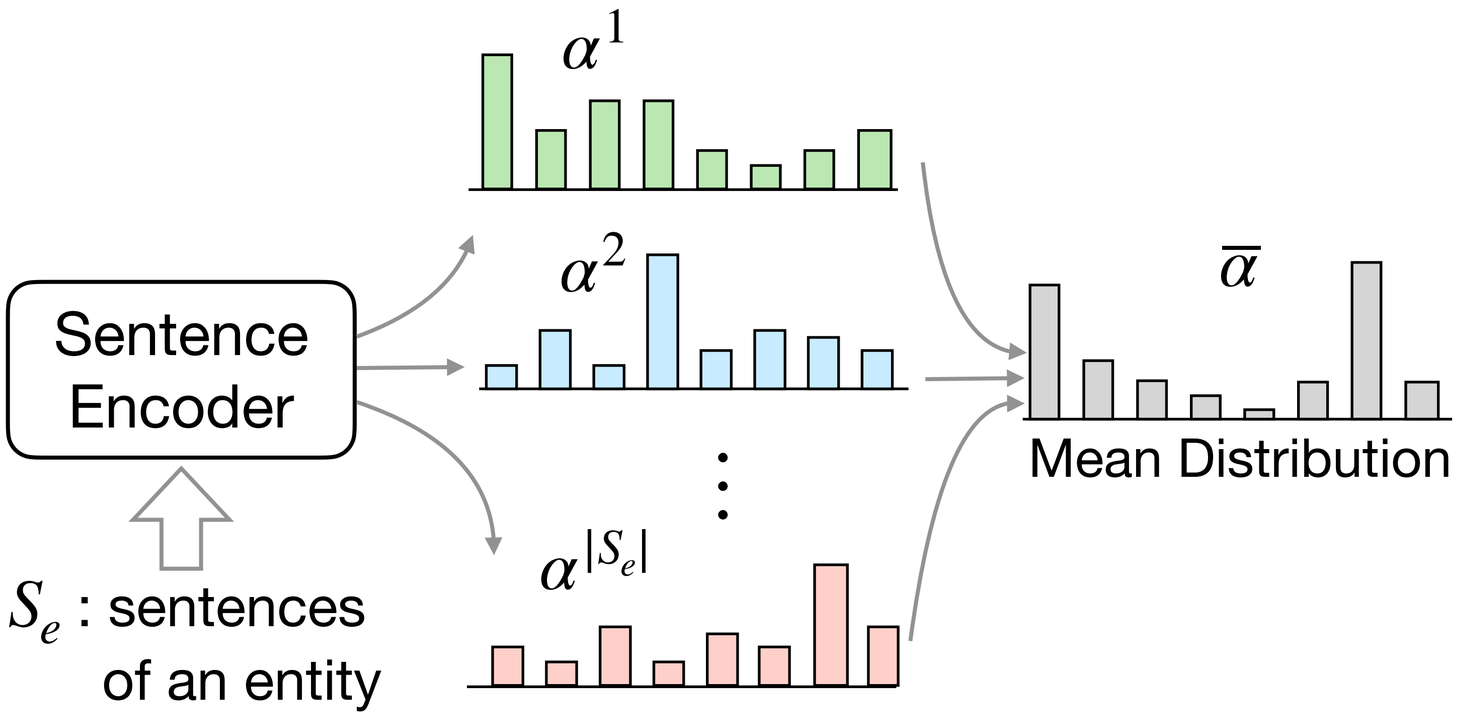
Summarization using α_h



Saliency score $\mathcal{R}(\alpha^s)$:

- **Relevance:** $\Delta(\bar{\alpha}, \alpha^s)$
- **Redundancy:** $-\gamma \max_{s' \in \hat{O}_e} \Delta(\alpha^{s'}, \alpha^s)$
- **Aspect-awareness:** Iterate over aspects and select salient sentences
- **Informativeness:** $-\beta\Delta(\alpha^B, \alpha^s)$

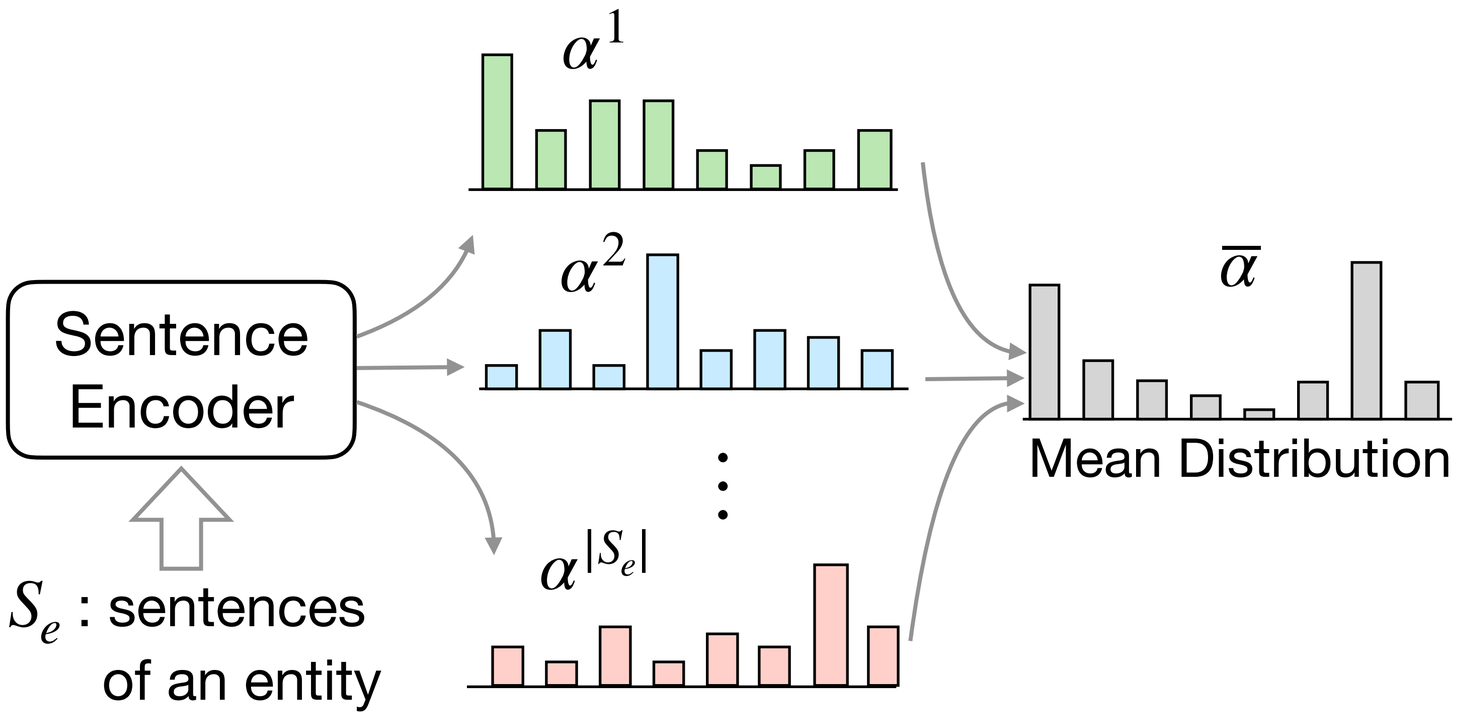
Summarization using α_h



Saliency score $\mathcal{R}(\alpha^s)$:

- **Relevance:** $\Delta(\bar{\alpha}, \alpha^s)$
- **Redundancy:** $-\gamma \max_{s' \in \hat{O}_e} \Delta(\alpha^{s'}, \alpha^s)$
- **Aspect-awareness:** Iterate over aspects and select salient sentences
- **Informativeness:** $-\beta\Delta(\alpha^B, \alpha^s)$

Summarization using α_h



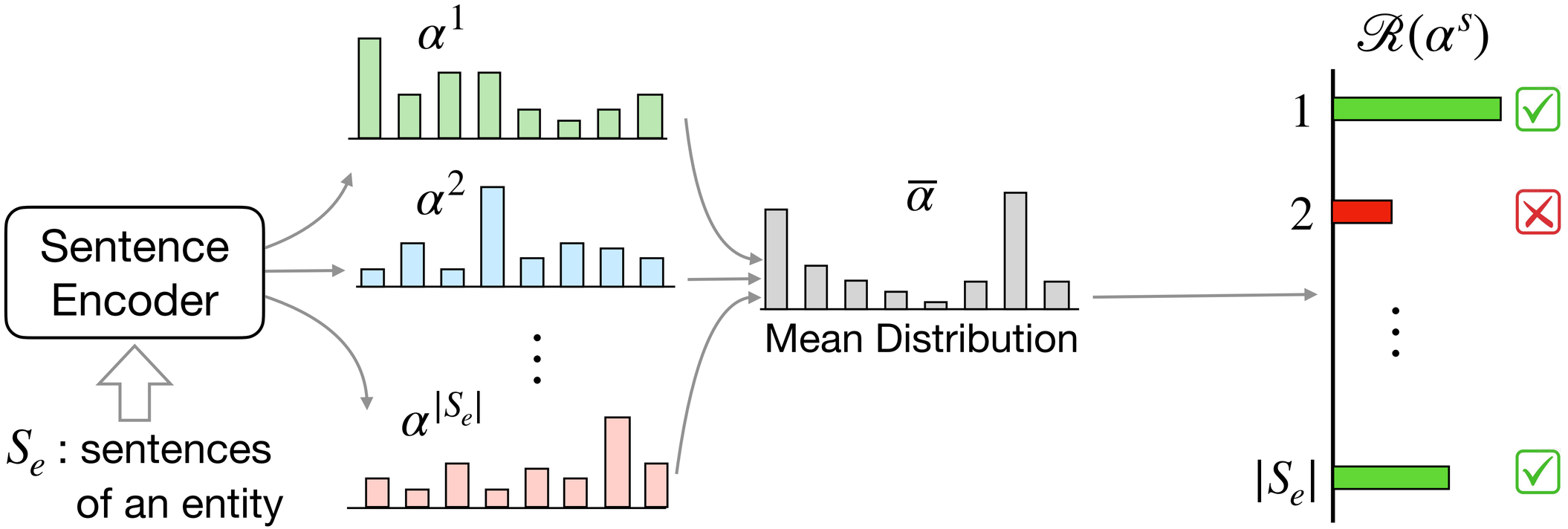
General Summarization:

$\mathcal{R}(\alpha^s)$ = [Relevance] - [Redundancy] + [Aspect-awareness]

Aspect Summarization:

$\mathcal{R}(\alpha^s)$ = [Relevance] + [Informativeness]

Summarization using α_h

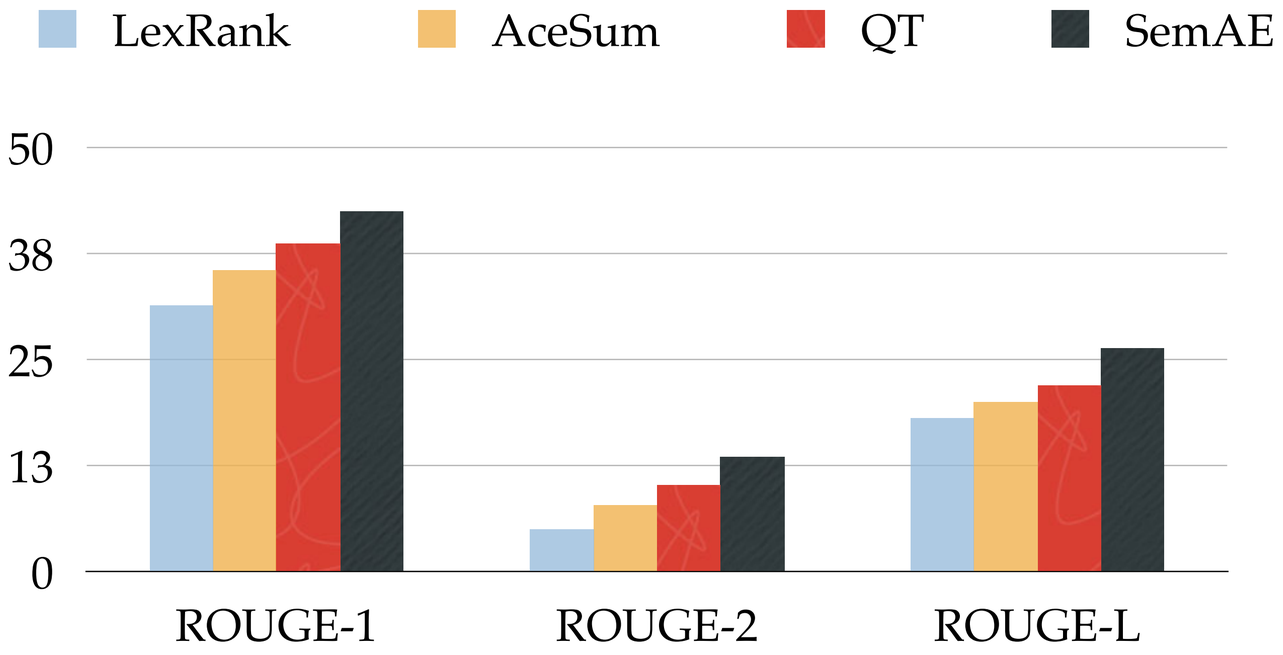


Experiments

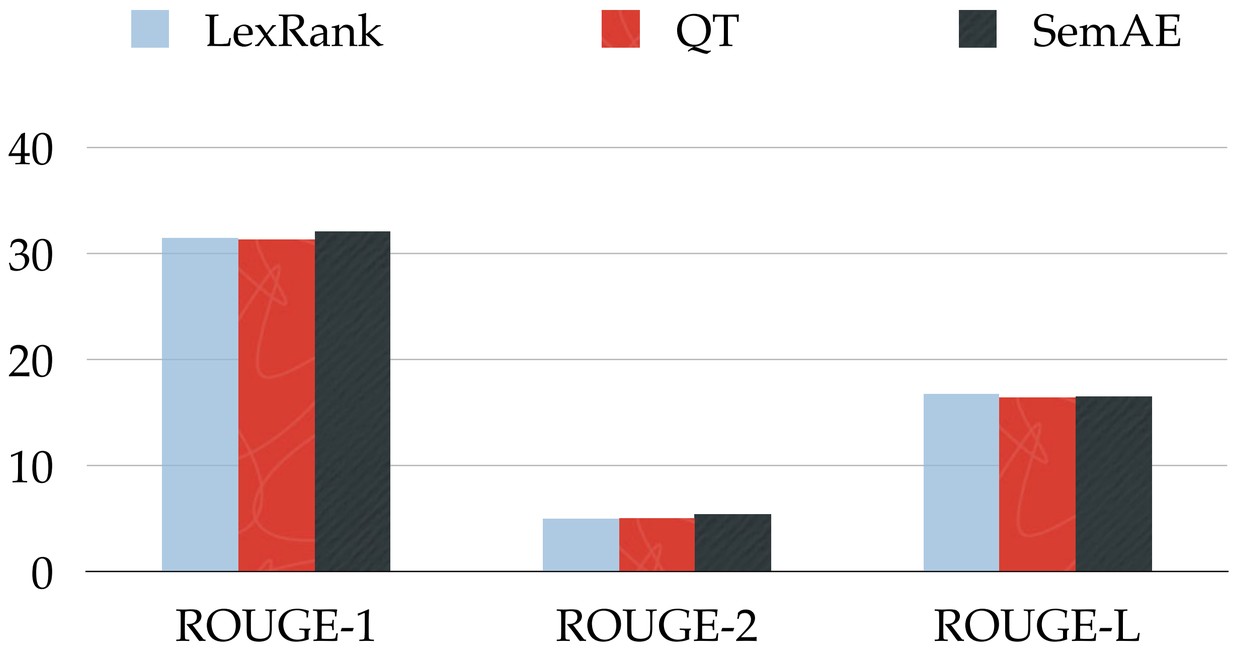
Datasets

- SPACE (Angelidis et al. 2021) — hotel reviews from Tripadvisor
- Amazon reviews (He and McAuley, 2016) — product reviews from amazon

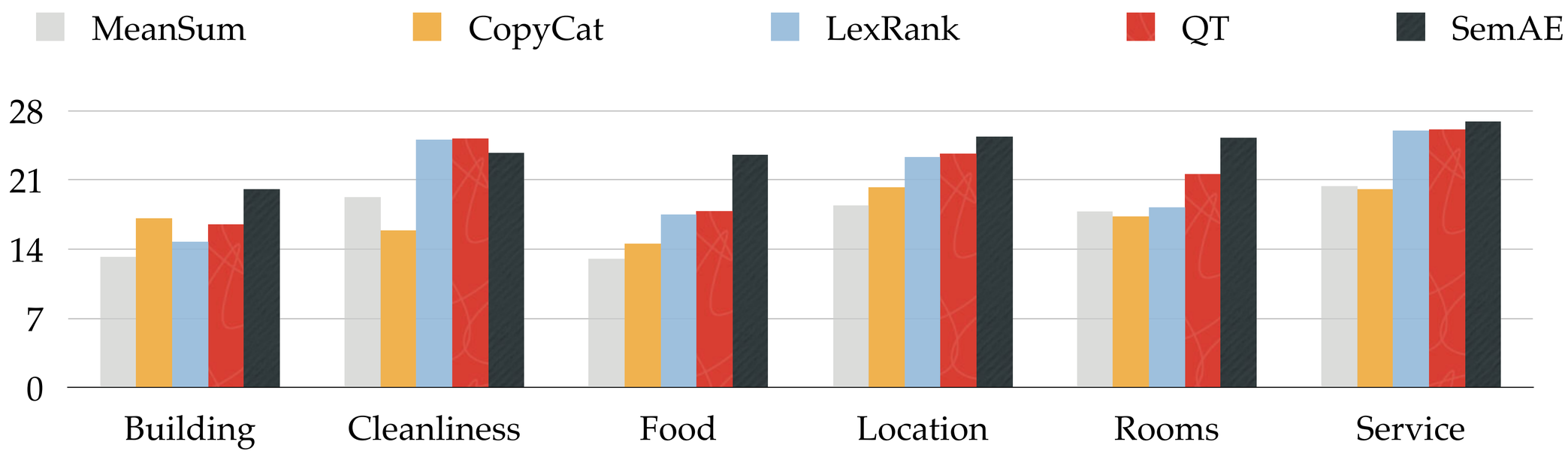
General Summarization - SPACE



General Summarization - Amazon



Aspect Summarization - SPACE



Summaries

SemAE

The staff is great. The Hotel Erwin is a great place to stay. **The staff were friendly and helpful. The location is perfect**. We ate breakfast at the hotel and it was great. **The hotel itself is in a great location. The service was wonderful**. It was great. The rooms are great.

Summaries

SemAE

The staff is great. The Hotel Erwin is a great place to stay. **The staff were friendly and helpful. The location is perfect**. We ate breakfast at the hotel and it was great. **The hotel itself is in a great location. The service was wonderful**. It was great. The rooms are great.

SemAE (w/ redun.)

The hotel itself is in a great **location**. The **rooms** were clean and we were on the 5th. The best part of the **hotel** is the 7th floor rooftop deck. The **staff** is great. The **hotel** has so many advantages over the other options in the area that it is a no contest.

Summaries

SemAE

The staff is great. The Hotel Erwin is a great place to stay. **The staff were friendly and helpful. The location is perfect**. We ate breakfast at the hotel and it was great. **The hotel itself is in a great location. The service was wonderful**. It was great. The rooms are great.

SemAE (w/ redun.)

The hotel itself is in a great location. The rooms were clean and we were on the 5th. The best part of the hotel is the 7th floor rooftop deck. The staff is great. The **hotel** has so many advantages over the other options in the area that it is a no contest.

SemAE (w/ redun. + asp.)

The **staff** is great. We had a great **stay** at the Erwin, and the staff really made it more enjoyable. The Hotel Erwin is a **great place** to stay. It was great. We ate **breakfast** at the hotel and it was great. The **food and service** at the restaurant was awesome. The **rooms** are great.

Controllable Summarization

Controllable Summarization

Positive

Love the warm chocolate chips cookies and the service has always been outstanding. Excellent morning breakfasts and the airport shuttle runs every 15 minutes but we have made the 10 minute walk numerous times to the airport terminal.

Controllable Summarization

Positive

Love the warm chocolate chips cookies and the service has always been outstanding. Excellent morning breakfasts and the airport shuttle runs every 15 minutes but we have made the 10 minute walk numerous times to the airport terminal.

Negative

To add insult to injury, for people who use the parking lot to "park and fly", the charge is $7.95/day, almost half of what the hotel guests are charged!! **Cons - Hotel is spread out so pay attention to how to get to your room as you may get lost.**

Controllable Summarization

Positive

Love the warm chocolate chips cookies and the service has always been outstanding. Excellent morning breakfasts and the airport shuttle runs every 15 minutes but we have made the 10 minute walk numerous times to the airport terminal.

Negative

To add insult to injury, for people who use the parking lot to "park and fly", the charge is $7.95/day, almost half of what the hotel guests are charged!! **Cons - Hotel is spread out so pay attention to how to get to your room as you may get lost.**

Neutral

Stayed at this hotel because the park n fly. **We have stayed at this hotel several times in the family suite (2 bedrooms/1 king and 2 queen beds).** Despite the enormity of this hotel, it very much feels almost family run.

Takeaways

- SemAE learns representations as distributions over latent units

Takeaways

- SemAE learns representations as distributions over latent units
- Sentence selection is performed using information-theoretic metrics

Takeaways

- SemAE learns representations as distributions over latent units
- Sentence selection is performed using information-theoretic metrics
- SemAE achieves strong performance on SPACE and Amazon

Takeaways

- SemAE learns representations as distributions over latent units
- Sentence selection is performed using information-theoretic metrics
- SemAE achieves strong performance on SPACE and Amazon
- SemAE can perform different forms of controllable summarization

Takeaways

- SemAE learns representations as distributions over latent units
- Sentence selection is performed using information-theoretic metrics
- SemAE achieves strong performance on SPACE and Amazon
- SemAE can perform different forms of controllable summarization

@SomnathBrc

brcsomnath/SemAE

somnath@cs.unc.edu

Comparing Summaries

SemAE

All staff members were friendly, accommodating, and helpful. The hotel and room were very clean. The room had modern charm and was nicely remodeled. The beds are extremely comfortable. The rooms are quite with wonderful beach views. The food at Hash, the restaurant in lobby, was fabulous. The location is great, very close to the beach. It's a longish walk to Santa Monica. The price is very affordable.

QT

The staff is great. The Hotel Erwin is a great place to stay. The staff were friendly and helpful. The location is perfect. We ate breakfast at the hotel and it was great. The hotel itself is in a great location. The service was wonderful. It was great.The rooms are great. The rooftop bar HIGH was the icing on the cake. The food and service at the restaurant was awesome. The service was excellent.

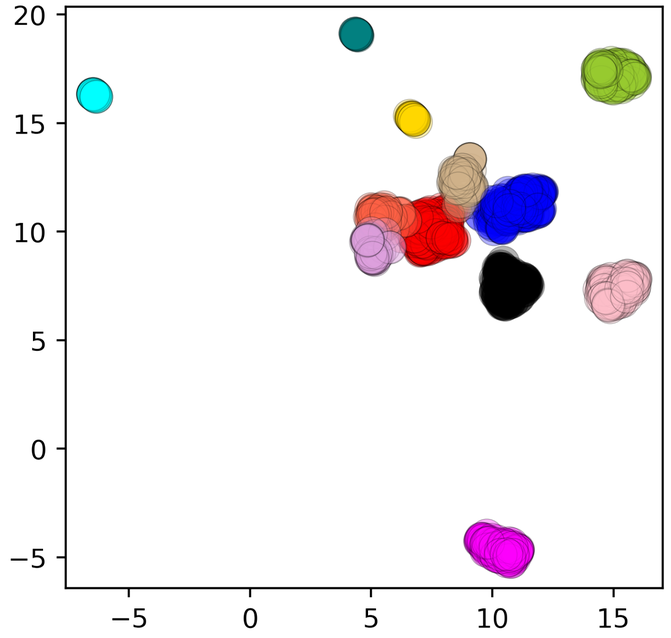
CopyCat

This hotel is in a great location, just off the beach. The staff was very friendly and helpful. We had a room with a view of the beach and ocean. The only problem was that our room was on the 4th floor with a view of the ocean. If you are looking for a nice place to sleep then this is the place for you.

Analysis - Dictionary Interpretation

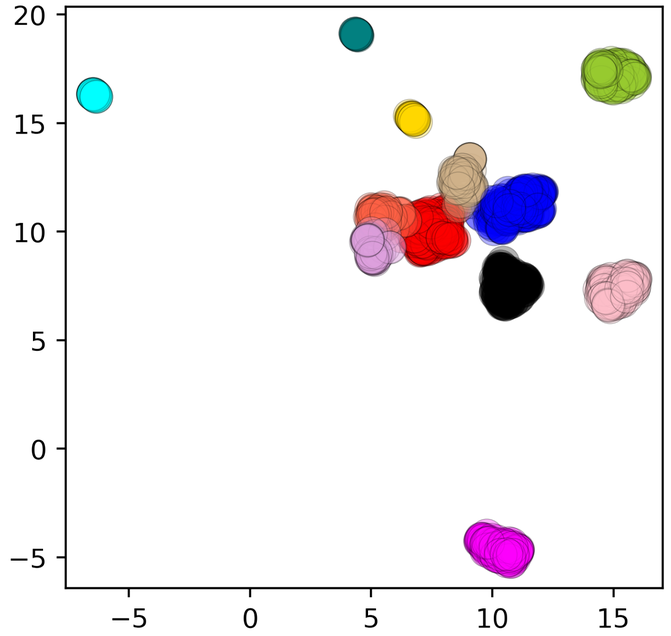
Analysis - Dictionary Interpretation

- Dictionary representations converges into clusters



Analysis - Dictionary Interpretation

- Dictionary representations converges into clusters
- Clusters capture distinct semantic meanings



Analysis - Dictionary Interpretation

Cluster #5

- I wish all hotels or any business for that matter, had **employees a dedicated to service** as he was.
- **Very polite and very professional** approach.

Good Service

Analysis - Dictionary Interpretation

Cluster #5

- I wish all hotels or any business for that matter, had **employees a dedicated to service** as he was.
- **Very polite and very professional** approach.

Good Service

Cluster #10

- **Stayed** here in August for the our first trip to Vancouver.
- I **stayed** at this motel with my partner in August 2010.

Phrase "Stayed"

Analysis - Dictionary Interpretation

Cluster #5

- I wish all hotels or any business for that matter, had **employees a dedicated to service** as he was.
- **Very polite and very professional** approach.

Good Service

Cluster #10

- **Stayed** here in August for the our first trip to Vancouver.
- I **stayed** at this motel with my partner in August 2010.

Phrase "Stayed"

Cluster #8

- **The rooms seem small**, tight fit for a family of 4.
- You may have a **difficult fit**.

Small rooms