Unsupervised Opinion Summarization Using Approximate Geodesics





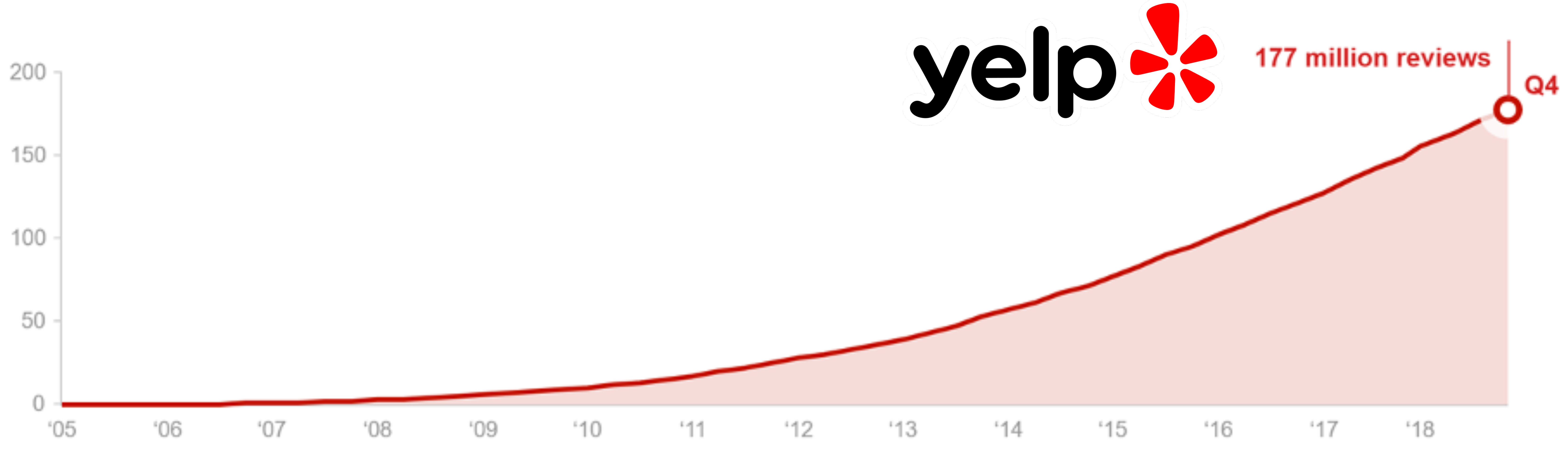
Somnath Basu Roy Chowdhury, Nicholas Monath, Avinava Dubey, Amr Ahmed, and Snigdha Chaturvedi

Outline

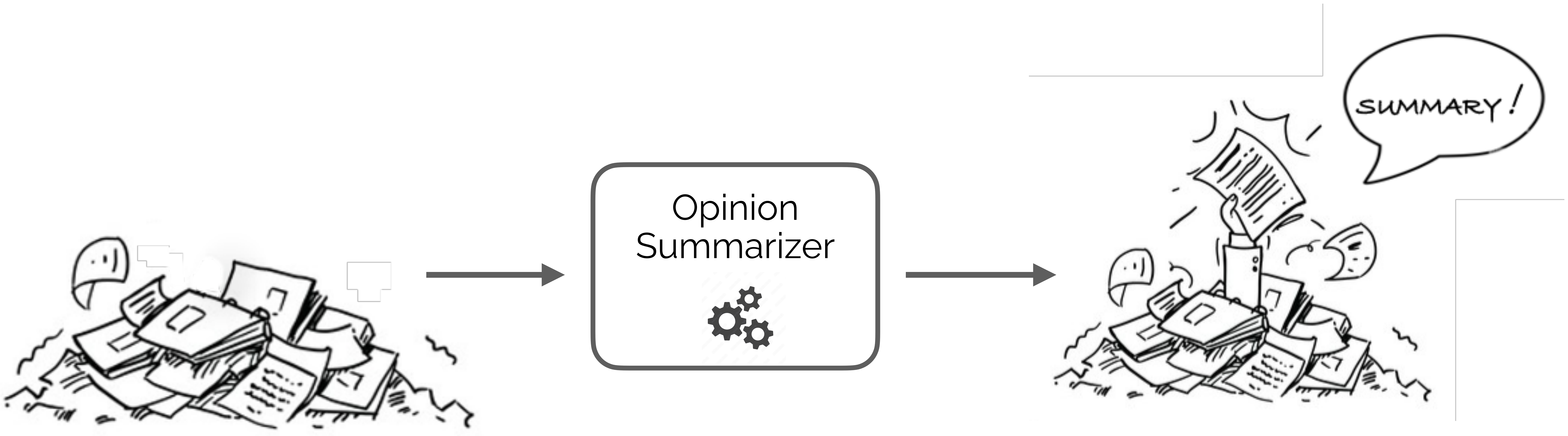
- Motivation
- Problem Setup
- Topical Representations
- Recipe for Summarization
- Representation Learning
- Sentence Selection
- Datasets and Metrics
- Results

Online Reviews

- There has a massive increase in the number of reviews available online
- These are a great resource for both sellers and customers



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 represent text in a way that they capture the underlying semantics

Motivation

- Unsupervised opinion systems are desirable due to the scarcity of labeled data
- It is crucial to represent text in a way that they capture the underlying semantics
- One such approach is to have representations as a distribution over latent semantic units

Motivation

- Unsupervised opinion systems are desirable due to the scarcity of labeled data
- It is crucial to represent text in a way that they capture the underlying semantics
- One such approach is to have representations as a distribution over latent semantic units
- Select popular opinions by leveraging such representations

Motivation

- Unsupervised opinion systems are desirable due to the scarcity of labeled data
- It is crucial to represent text in a way that they capture the underlying semantics
- One such approach is to have representations as a distribution over latent semantic units
- Select popular opinions by leveraging such representations
- We focus on extractive summarization in this work

Outline

- Motivation
- Problem Setup
- Topical Representations
- Recipe for Summarization
- Representation Learning
- Sentence Selection
- Datasets and Metrics
- Results

Problem Setup

- For an entity (a product — kindle, a hotel — Graduate CH), an opinion set is provided
- Extract a set of review sentences to form a summary
- Compare the generated summary with a human-written one

Outline

- Motivation
- Problem Setup
- Topical Representations
- Recipe for Summarization
- Representation Learning
- Sentence Selection
- Datasets and Metrics
- Results

Distributed Representations

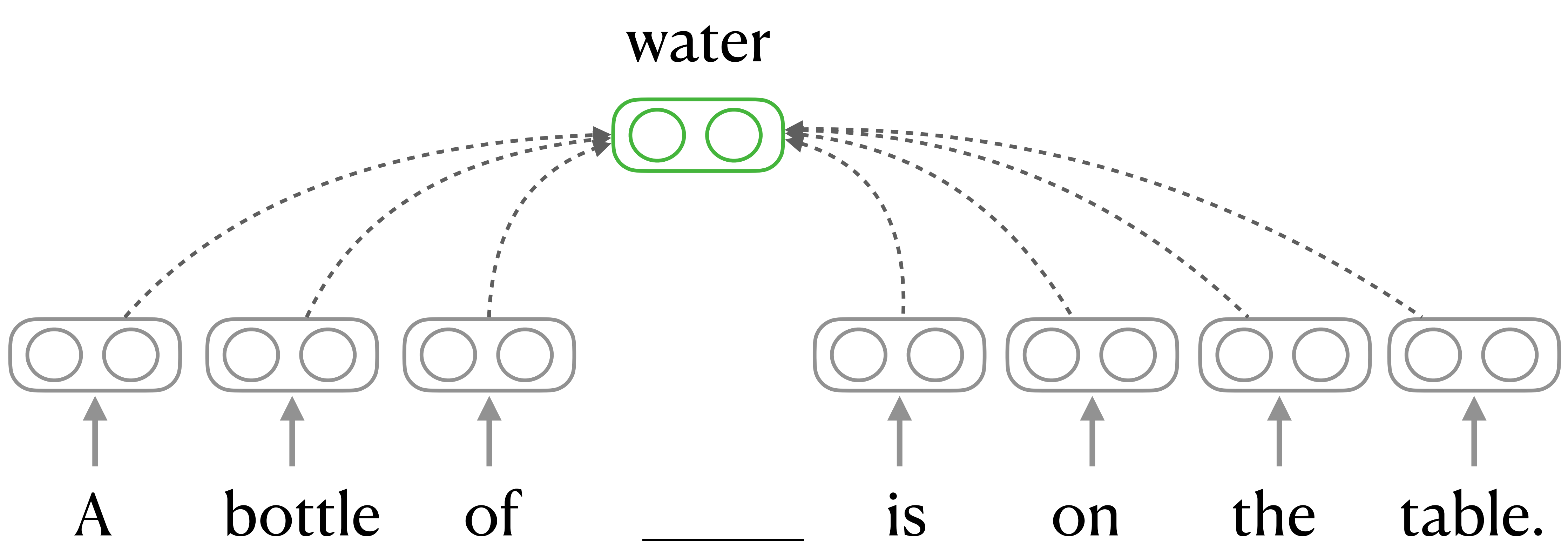
Distributional hypothesis (Firth et al. 1950):

"a word is characterized by the company it keeps"

Distributed Representations

Distributional hypothesis (Firth et al. 1950):

"a word is characterized by the company it keeps"

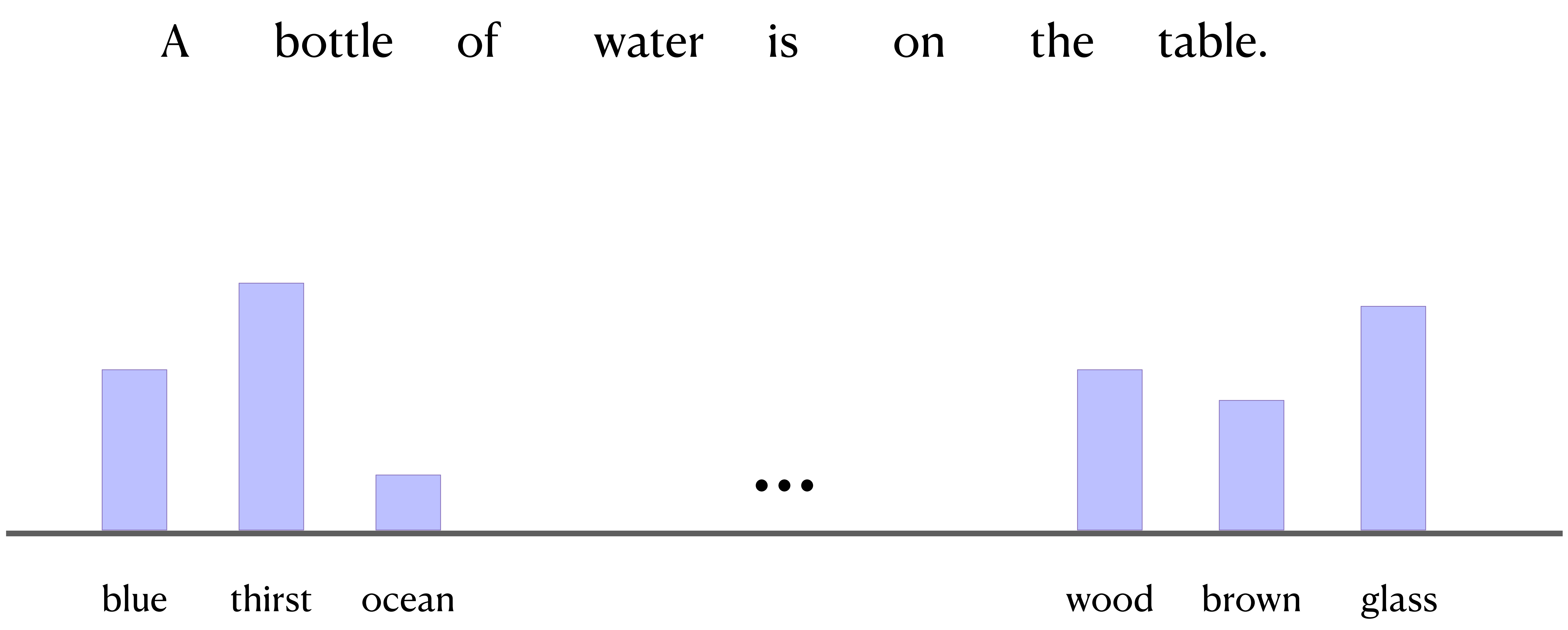


Topical Representations

We want to capture the meaning of text as a distribution over semantic units.

Topical Representations

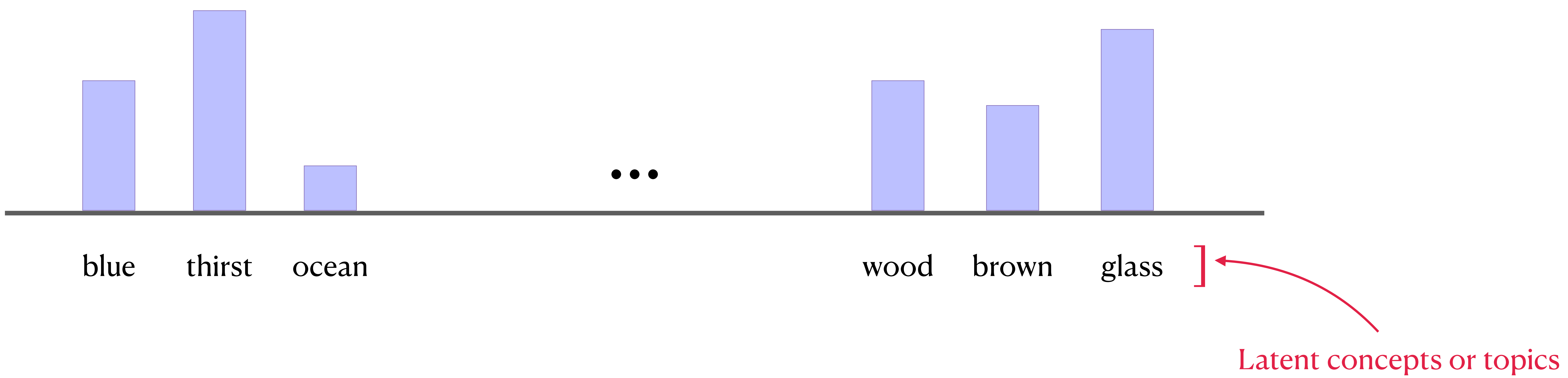
We want to capture the meaning of text as a distribution over semantic units.



Topical Representations

We want to capture the meaning of text as a distribution over semantic units.

A bottle of water is on the table.



Why not distributed representations?

- It is hard to use off-the-shelf pre-trained distributions for defining similarity measures

Why not distributed representations?

- It is hard to use off-the-shelf pre-trained distributions for defining similarity measures
- (Timkey et al, 2021) showed that BERT representations are anisotropic in nature

Why not distributed representations?

- It is hard to use off-the-shelf pre-trained distributions for defining similarity measures
- (Timkey et al, 2021) showed that BERT representations are anisotropic in nature
- Few dimensions dominate the similarity scores

Why not distributed representations?

- It is hard to use off-the-shelf pre-trained distributions for defining similarity measures
- (Timkey et al, 2021) showed that BERT representations are anisotropic in nature
- Few dimensions dominate the similarity scores
- Hard to achieve compositionality using standard operations (add or mul.)

Why topical representations?

- Representations are distributions over the same support

Why topical representations?

- Representations are distributions over the same support
- Allows us to compare representations using cosine similarity

Why topical representations?

- Representations are distributions over the same support
- Allows us to compare representations using cosine similarity
- Retrieve overall semantic distribution using an aggregate (mean) representation

Outline

- Motivation
- Problem Setup
- Topical Representations
- Recipe for Summarization
- Representation Learning
- Sentence Selection
- Datasets and Metrics
- Results

Recipe for Summarization?

Our approach has 2 components:

Recipe for Summarization?

Our approach has 2 components:

- Unsupervised Representation Learning
  - Converts distributed representations → topical representations

Recipe for Summarization?

Our approach has 2 components:

- Unsupervised Representation Learning
    - Converts distributed representations → topical representations
- Sentence selection algorithm
    - Use topical representations to quantify relevance of a review sentence

Outline

- Motivation
- Problem Setup
- Topical Representations
- Recipe for Summarization
- Representation Learning
- Sentence Selection
- Datasets and Metrics
- Results

Unsupervised Representation Learning

- We use dictionary learning to decompose pre-trained representations into topical representations

Unsupervised Representation Learning

- We use dictionary learning to decompose pre-trained representations into topical representations

- The dictionary captures latent semantic units

Unsupervised Representation Learning

- We use dictionary learning to decompose pre-trained representations into topical representations
- The dictionary captures latent semantic units
- The sparse coefficients function as the topical representation

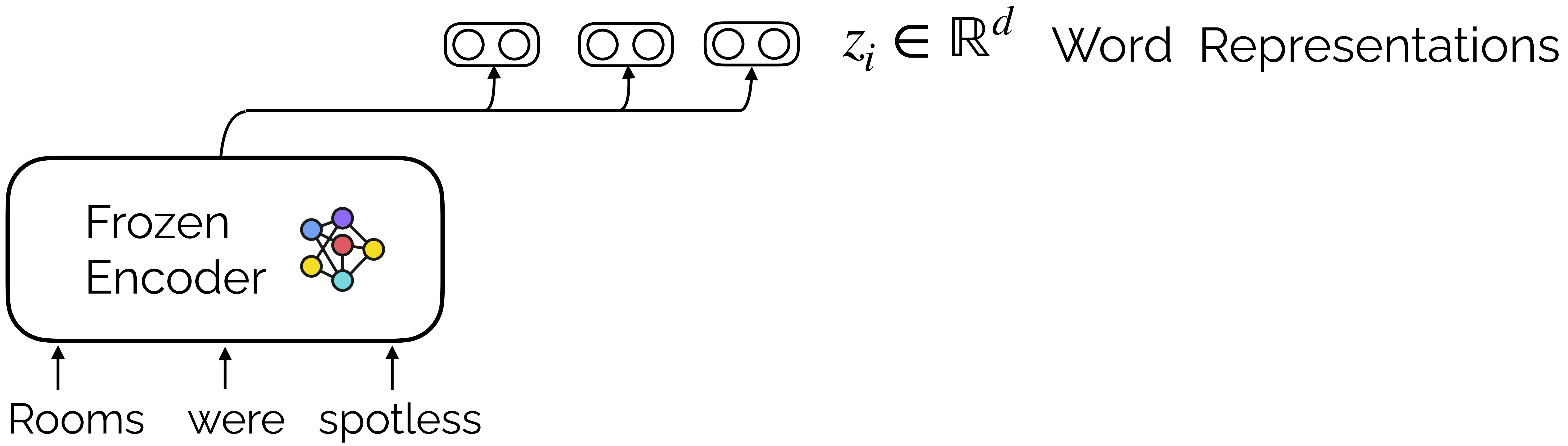
Unsupervised Representation Learning

- We use dictionary learning to decompose pre-trained representations into topical representations
- The dictionary captures latent semantic units
- The sparse coefficients function as the topical representation
- We use a sentence reconstruction objective for learning the dictionary

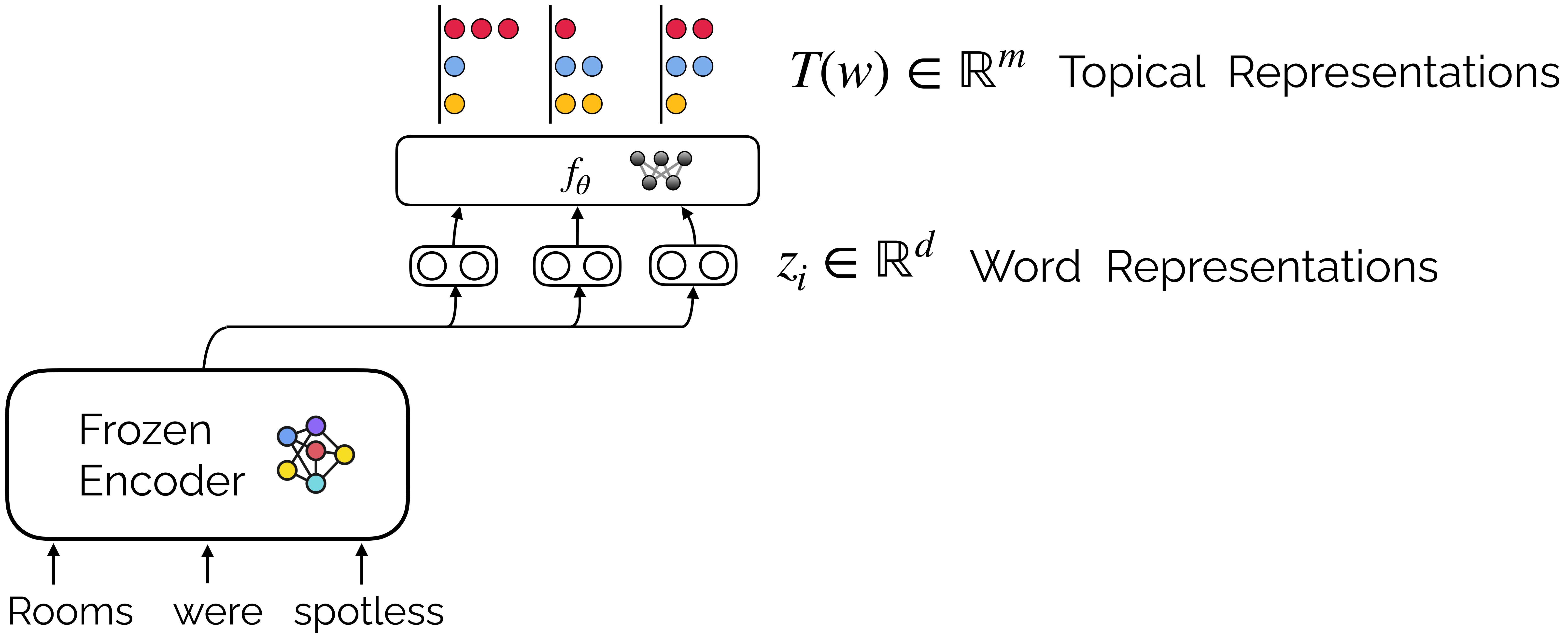
Unsupervised Representation Learning

- We use dictionary learning to decompose pre-trained representations into topical representations
- The dictionary captures latent semantic units
- The sparse coefficients function as the topical representation
- We use a sentence reconstruction objective for learning the dictionary
- We design an **encoder-decoder architecture** to achieve this

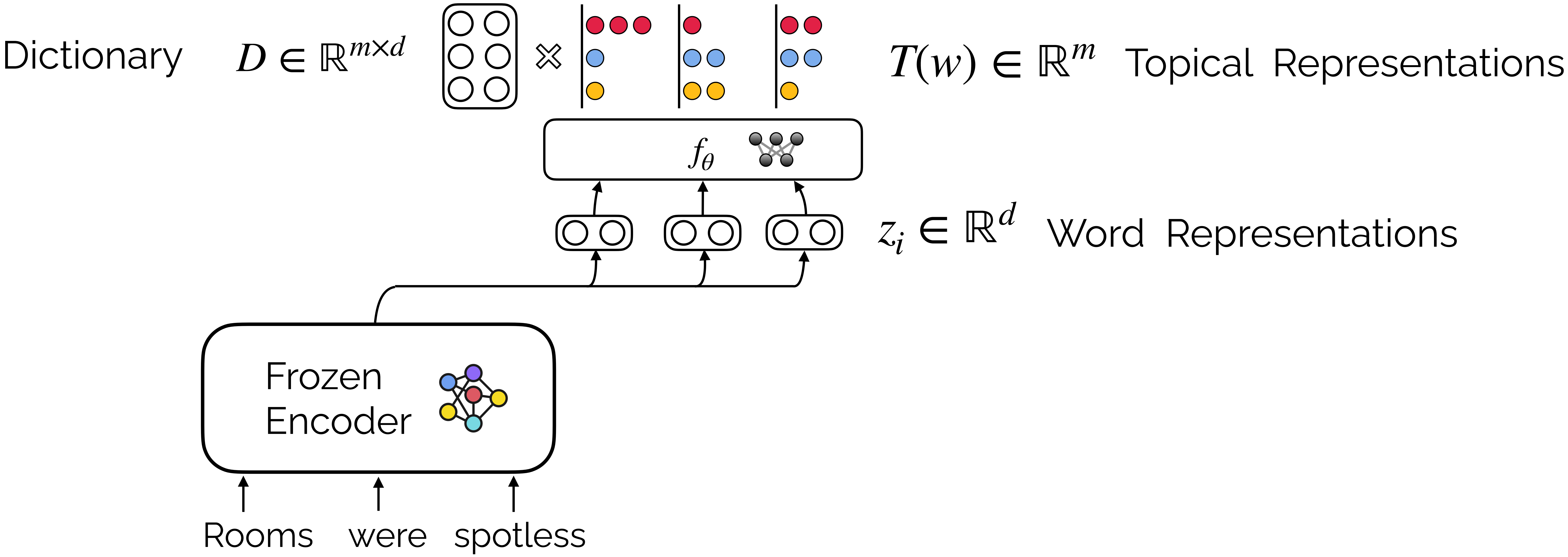
Model Sketch



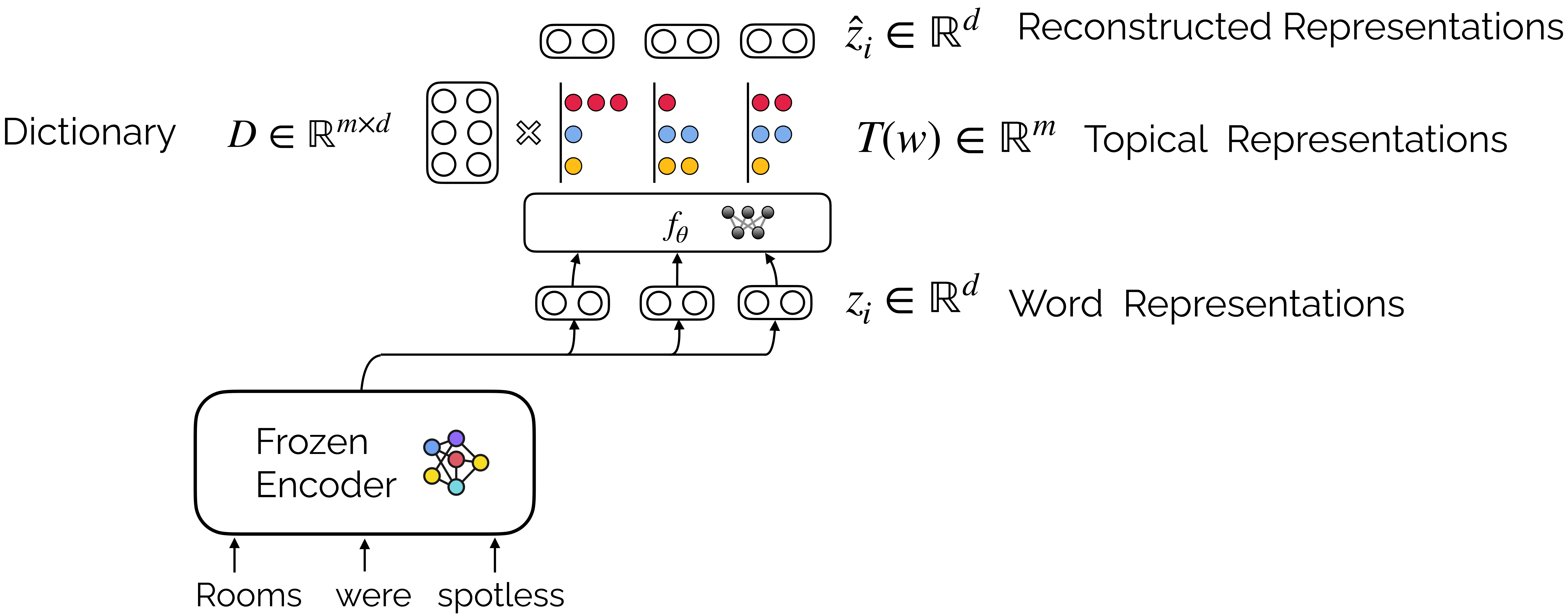
Model Sketch



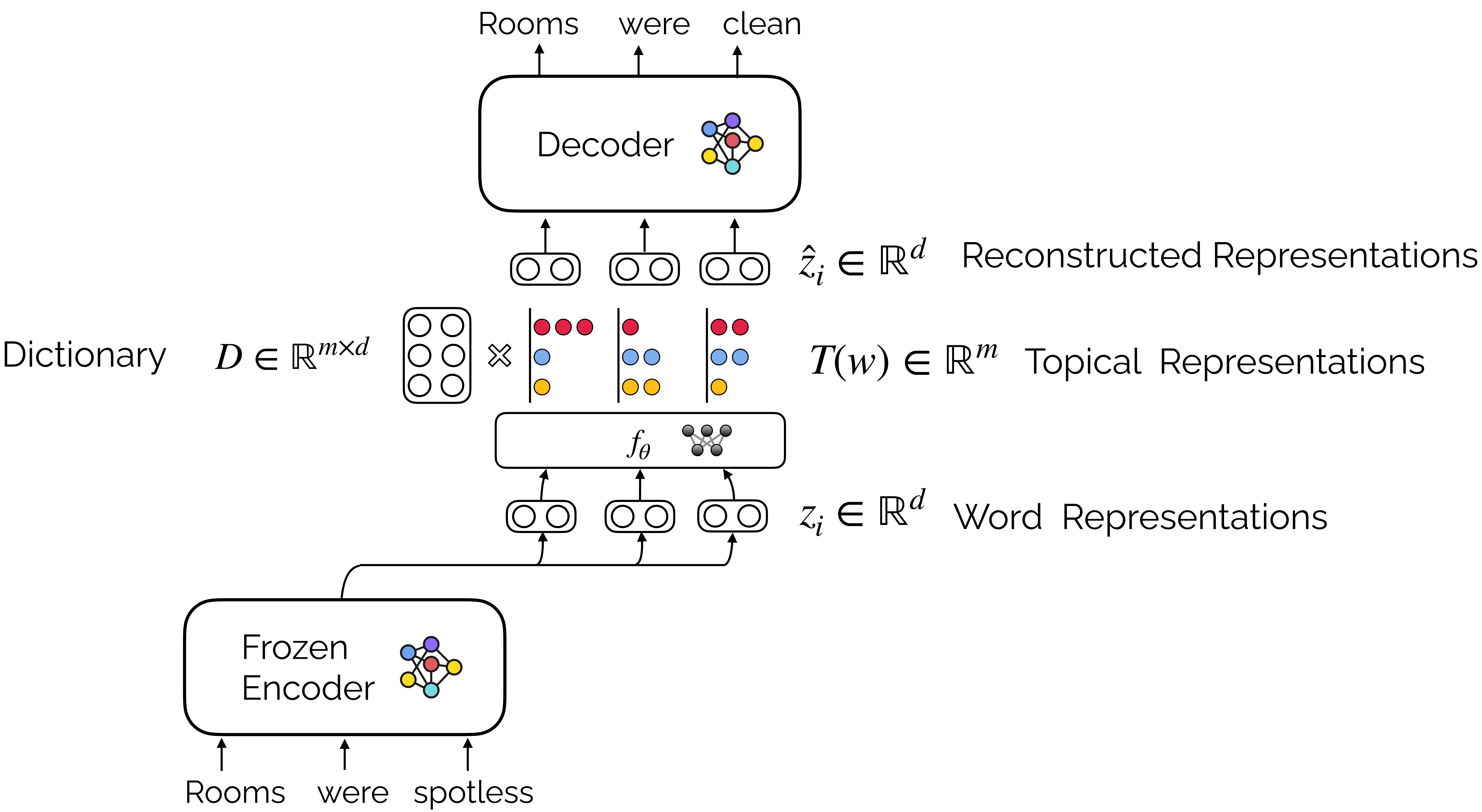
Model Sketch



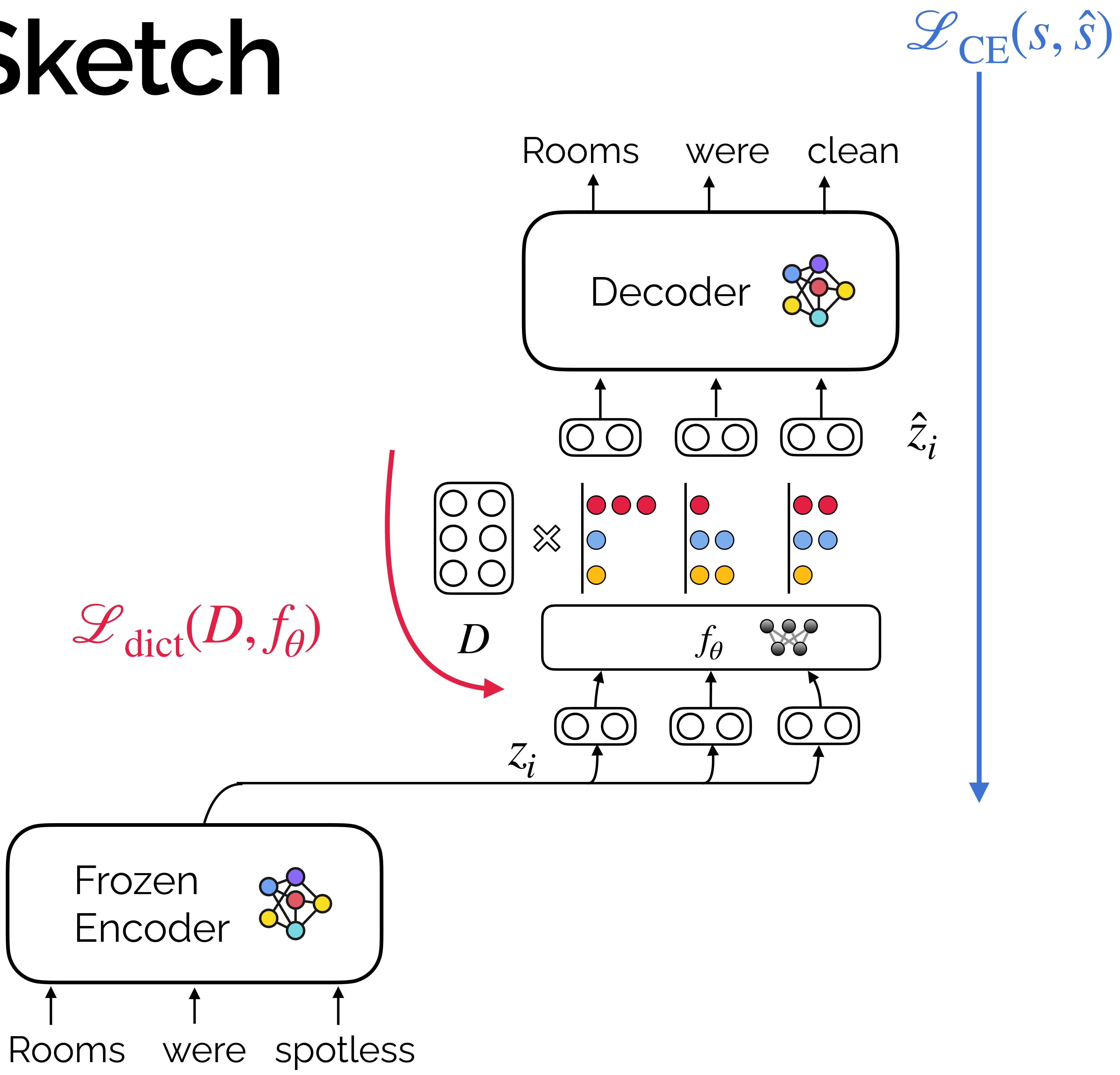
Model Sketch



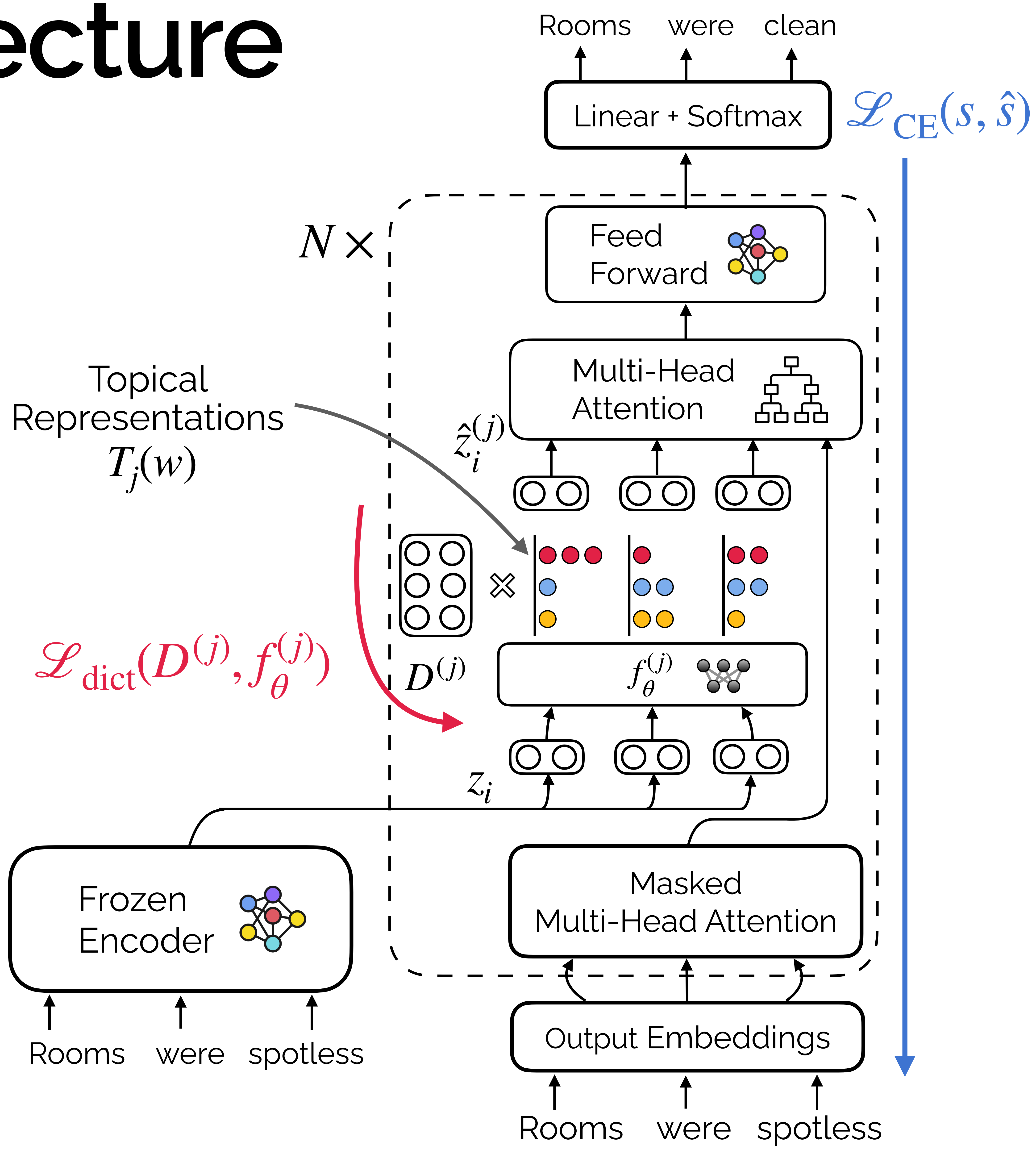
Model Sketch



Model Sketch



Model Architecture



Take aways

- We train our model using a combination of dictionary and cross-entropy loss

Take aways

- We train our model using a combination of dictionary and cross-entropy loss
- We maintain a separate dictionary at each decoder layer

Take aways

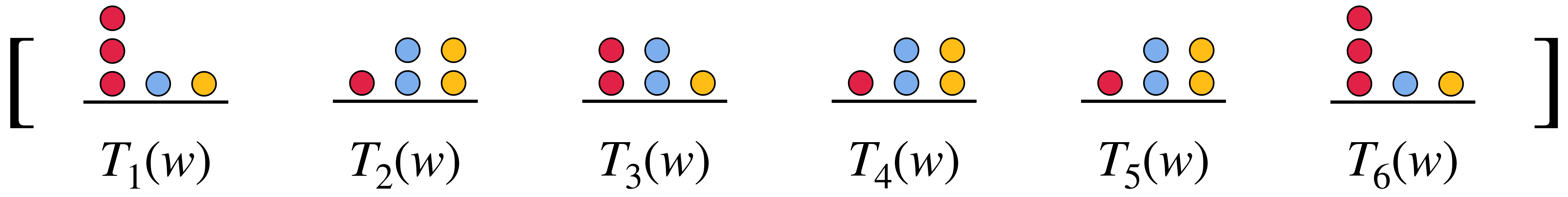
- We train our model using a combination of dictionary and cross-entropy loss
- We maintain a separate dictionary at each decoder layer
- We obtain a word representation for each decoder layer

Take aways

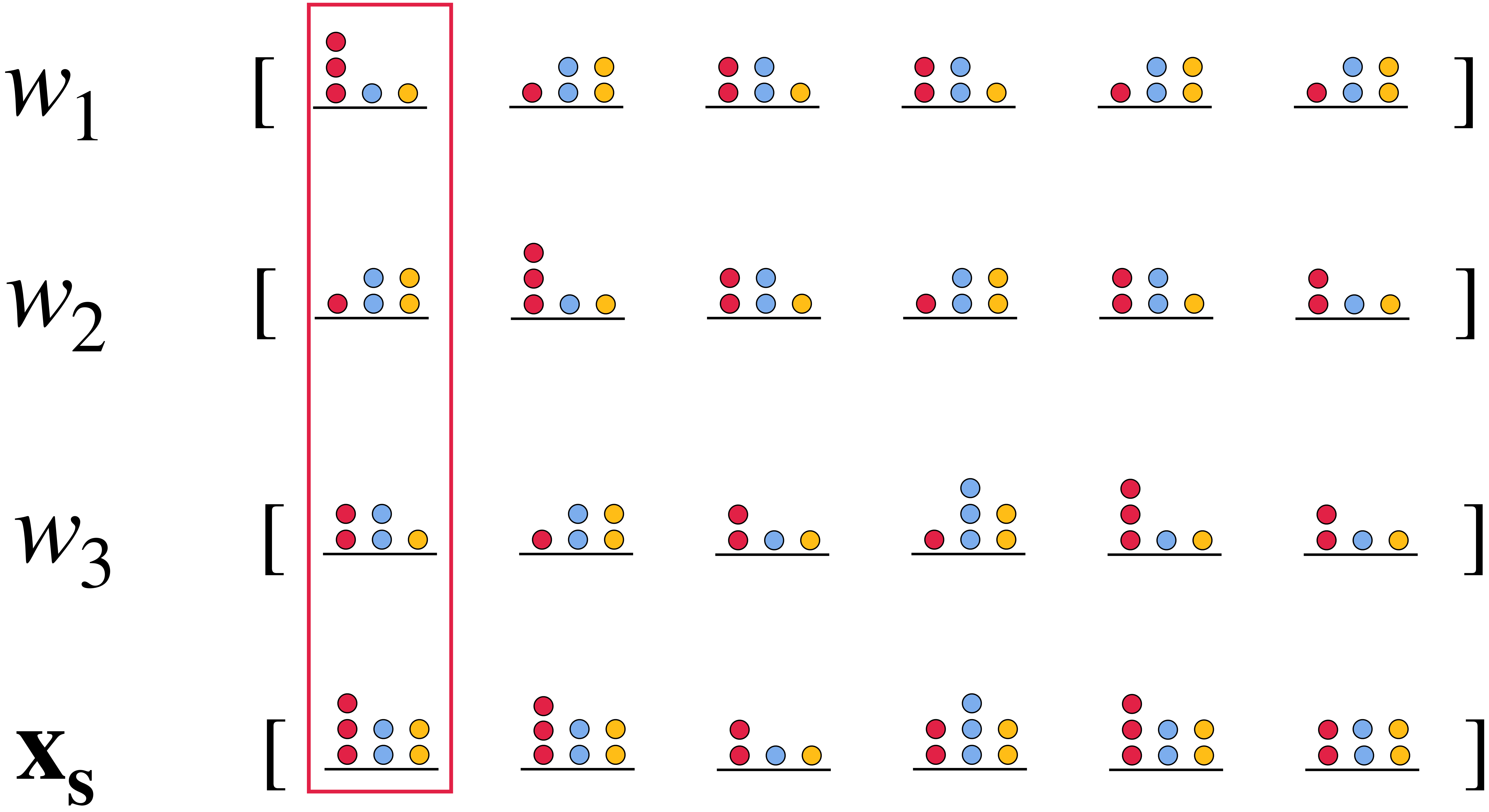
- We train our model using a combination of dictionary and cross-entropy loss
- We maintain a separate dictionary at each decoder layer
- We obtain a word representation for each decoder layer
- How do we combine these to form a sentence representation?

Word → Sentence Representations

w



Word → Sentence Representations



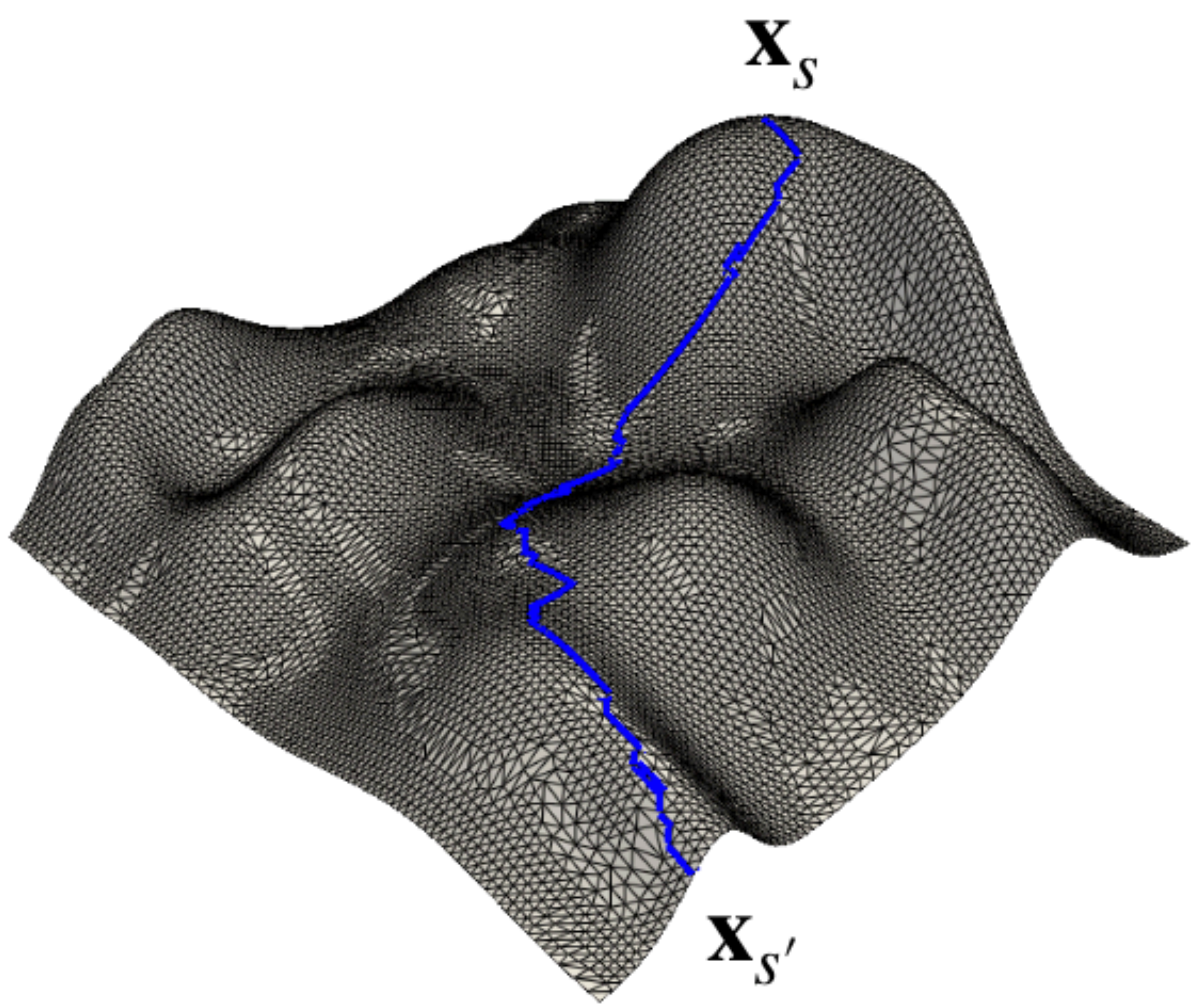
Outline

- Motivation
- Problem Setup
- Topical Representations
- Recipe for Summarization
- Representation Learning
- Sentence Selection
- Datasets and Metrics
- Results

Sentence Selection

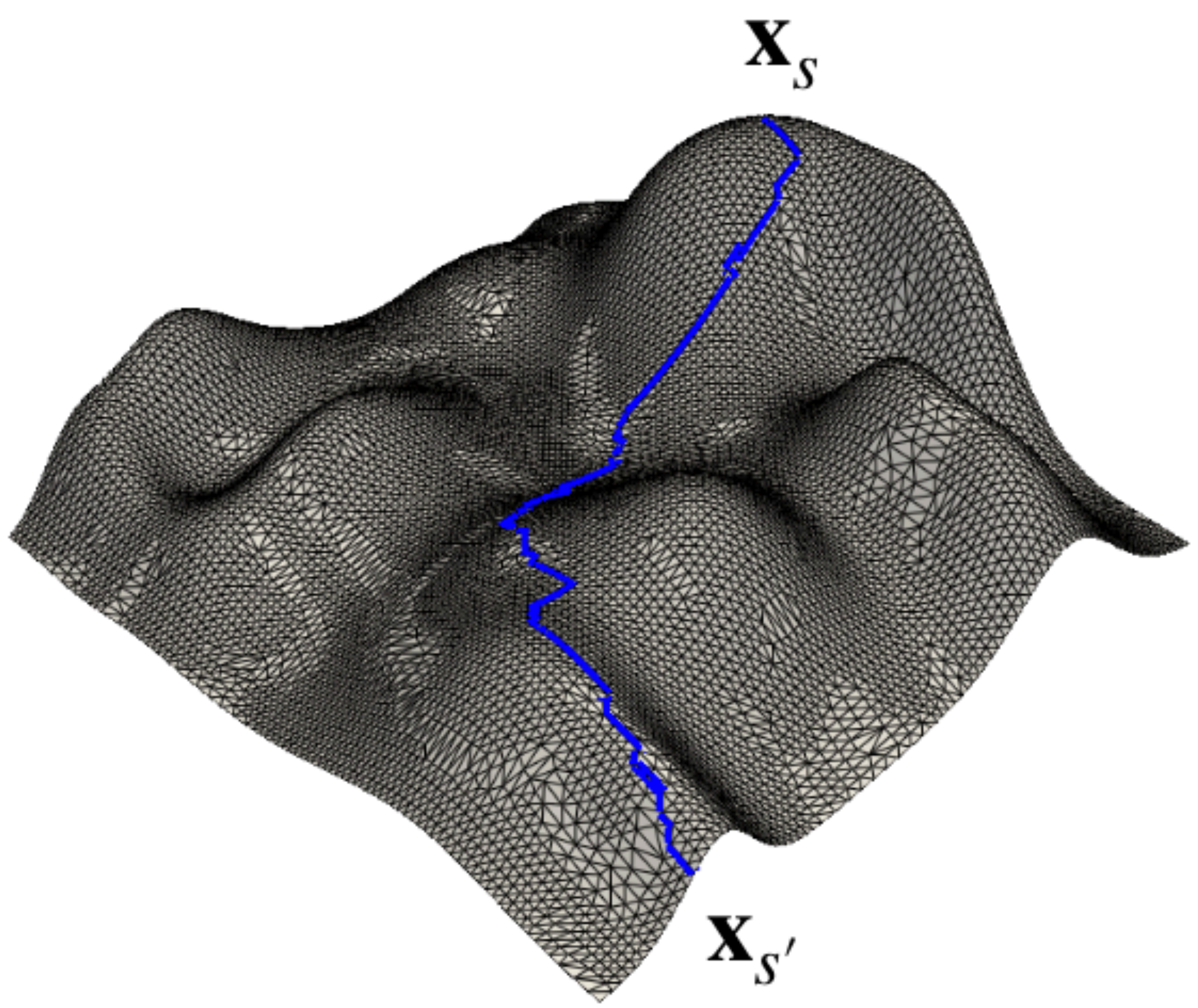
- We want to select sentences that are representative of popular opinions
- For an entity e, we have sentence representations $X_e = \{\mathbf{x}_\mathbf{S} \;\; s \in S_e\}$
- A naive approach is to select sentence representations close to the mean of X_e
- Sentence representations lie on a high-dimensional manifold that we need to consider while computing distances

Sentence Selection



Geodesic distance between two representations $\mathbf{x}_S$ and $\mathbf{x}_{S'}$

Sentence Selection



We name our system **Geo**desic **Summ**arizer (GeoSumm)

Sentence Selection

Algorithm 1 General Summarization Routine

Input: A set of sentence representations $\mathcal{X}_e = \{\mathbf{x}_s | s \in \mathcal{S}_e\}$ are review sentences for entity e.
$\mu_e \leftarrow \mathbb{E}_{s \sim \mathcal{S}_e}[\mathbf{x}_s]$
$\mathbf{A} \leftarrow \text{knn}(\mathcal{X}_e \cup \mu_e) \in \mathbb{R}^{l \times l}$ ▷ adjacency matrix of k-NN graph, $l = |S_e| + 1$.
$d \leftarrow \text{Dijkstra}(\mathbf{A}, \mu_e)$ ▷ shortest distances of all nodes from μ_e
$I = \{1/d(s) | s \in \mathcal{S}_e\}$ ▷ importance scores
$t_q \leftarrow \min \text{top-}q(I)$ ▷ top-q threshold
$\mathcal{O}_e \leftarrow \{s \mid I(s) \geq t_q, s \in \mathcal{S}_e\}$
return $\mathcal{O}_e$

Approximates the manifold structure using a kNN graph

Sentence Selection

Algorithm 1 General Summarization Routine

Input: A set of sentence representations $\mathcal{X}_e = \{\mathbf{x}_s | s \in \mathcal{S}_e\}$ are review sentences for entity e.
$\mu_e \leftarrow \mathbb{E}_{s \sim \mathcal{S}_e}[\mathbf{x}_s]$
$\mathbf{A} \leftarrow \mathrm{knn}(\mathcal{X}_e \cup \mu_e) \in \mathbb{R}^{l \times l}$ ▷ adjacency matrix of k-NN graph, $l = |S_e| + 1$.
$d \leftarrow \mathrm{Dijkstra}(\mathbf{A}, \mu_e)$ ▷ shortest distances of all nodes from μ_e
$I = \{1/d(s) | s \in \mathcal{S}_e\}$ ▷ importance scores
$t_q \leftarrow \min \text{top-}q(I)$ ▷ top-q threshold
$\mathcal{O}_e \leftarrow \{s \mid I(s) \geq t_q, s \in \mathcal{S}_e\}$
return $\mathcal{O}_e$

Computing the distances along the manifold

Sentence Selection

Algorithm 1 General Summarization Routine

Input: A set of sentence representations $\mathcal{X}_e = \{\mathbf{x}_s | s \in \mathcal{S}_e\}$ are review sentences for entity e.

$\mu_e \leftarrow \mathbb{E}_{s \sim \mathcal{S}_e}[\mathbf{x}_s]$

$\mathbf{A} \leftarrow \mathrm{knn}(\mathcal{X}_e \cup \mu_e) \in \mathbb{R}^{l \times l}$ ▷ adjacency matrix of k-NN graph, $l = |S_e| + 1$.

$d \leftarrow \mathrm{Dijkstra}(\mathbf{A}, \mu_e)$ ▷ shortest distances of all nodes from μ_e

$I = \{1/d(s) | s \in \mathcal{S}_e\}$ ▷ importance scores

$t_q \leftarrow \min \text{top-}q(I)$ ▷ top-q threshold

$\mathcal{O}_e \leftarrow \{s \mid I(s) \geq t_q, s \in \mathcal{S}_e\}$

return $\mathcal{O}_e$

Distances serve as the importance of a sentence

Sentence Selection

Algorithm 1 General Summarization Routine

1: **Input**: A set of sentence representations $\mathcal{X}_e = \{\mathbf{x}_s | s \in \mathcal{S}_e\}$ are review sentences for entity e.
2: $\mu_e \leftarrow \mathbb{E}_{s \sim \mathcal{S}_e}[\mathbf{x}_s]$
3: $\mathbf{A} \leftarrow \text{knn}(\mathcal{X}_e \cup \mu_e) \in \mathbb{R}^{l \times l}$ ▷ adjacency matrix of k-NN graph, $l = |S_e| + 1$.
4: $d \leftarrow \text{Dijkstra}(\mathbf{A}, \mu_e)$ ▷ shortest distances of all nodes from μ_e
5: $I = \{1/d(s) | s \in \mathcal{S}_e\}$ ▷ importance scores
6: $t_q \leftarrow \min \text{top-}q(I)$ ▷ top-q threshold
7: $\mathcal{O}_e \leftarrow \{s \mid I(s) \geq t_q, s \in \mathcal{S}_e\}$
8: **return** $\mathcal{O}_e$

Select top q sentences as the Output summary

Aspect Summarization

- Users often find aspect-specific summaries useful

Aspect Summarization

- Users often find aspect-specific summaries useful
- For example, a hotel entity has different aspects — food, rooms, service etc.

Aspect Summarization

- Users often find aspect-specific summaries useful
- For example, a hotel entity has different aspects — food, rooms, service etc.
- Our framework supports this by using an aspect-specific mean representation

Aspect Summarization

- Users often find aspect-specific summaries useful
- For example, a hotel entity has different aspects — food, rooms, service etc.
- Our framework supports this by using an aspect-specific mean representation
- Aspect sentences are identified using keywords provided in the dataset

Outline

- Motivation
- Problem Setup
- Topical Representations
- Recipe for Summarization
- Representation Learning
- Sentence Selection
- Datasets and Metrics
- Results

Dataset

- OPOSUM+ (Amplayo et al. 2021) — products reviews (e.g. laptops, bags) from amazon
- SPACE (Angelidis et al. 2021) — hotel reviews from Tripadvisor
- Amazon reviews (He and McAuley, 2016) — product reviews (e.g. electronics) from amazon

Evaluation

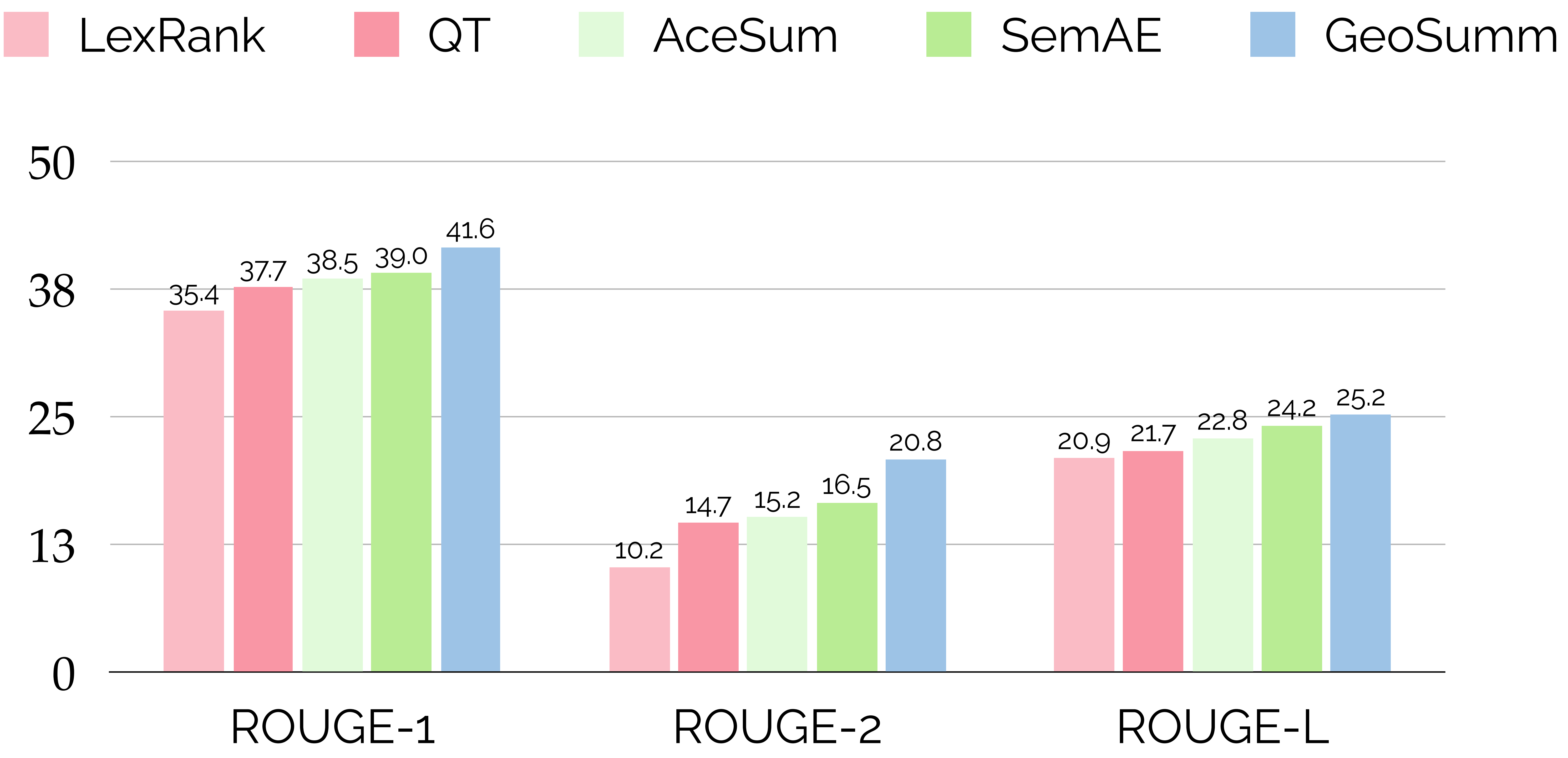
We compare the lexical overlap between system and reference summaries

- ROUGE-1 - refers to overlap of *unigrams* (words)
- ROUGE-2 - refers to overlap of *bigrams*
- ROUGE-L - considers the longest common subsequence

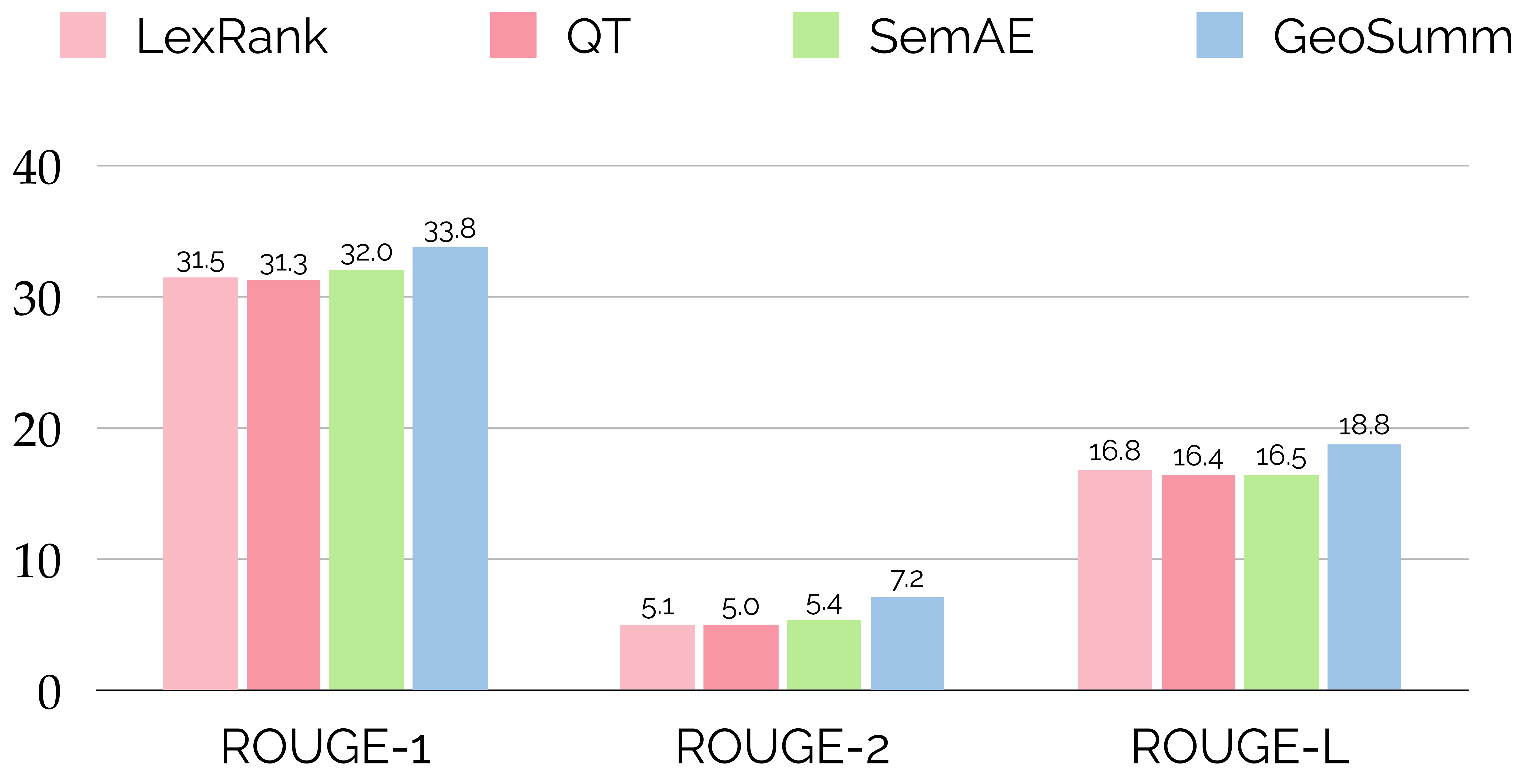
Outline

- Motivation
- Problem Setup
- Topical Representations
- Recipe for Summarization
- Representation Learning
- Sentence Selection
- Datasets and Metrics
- Results

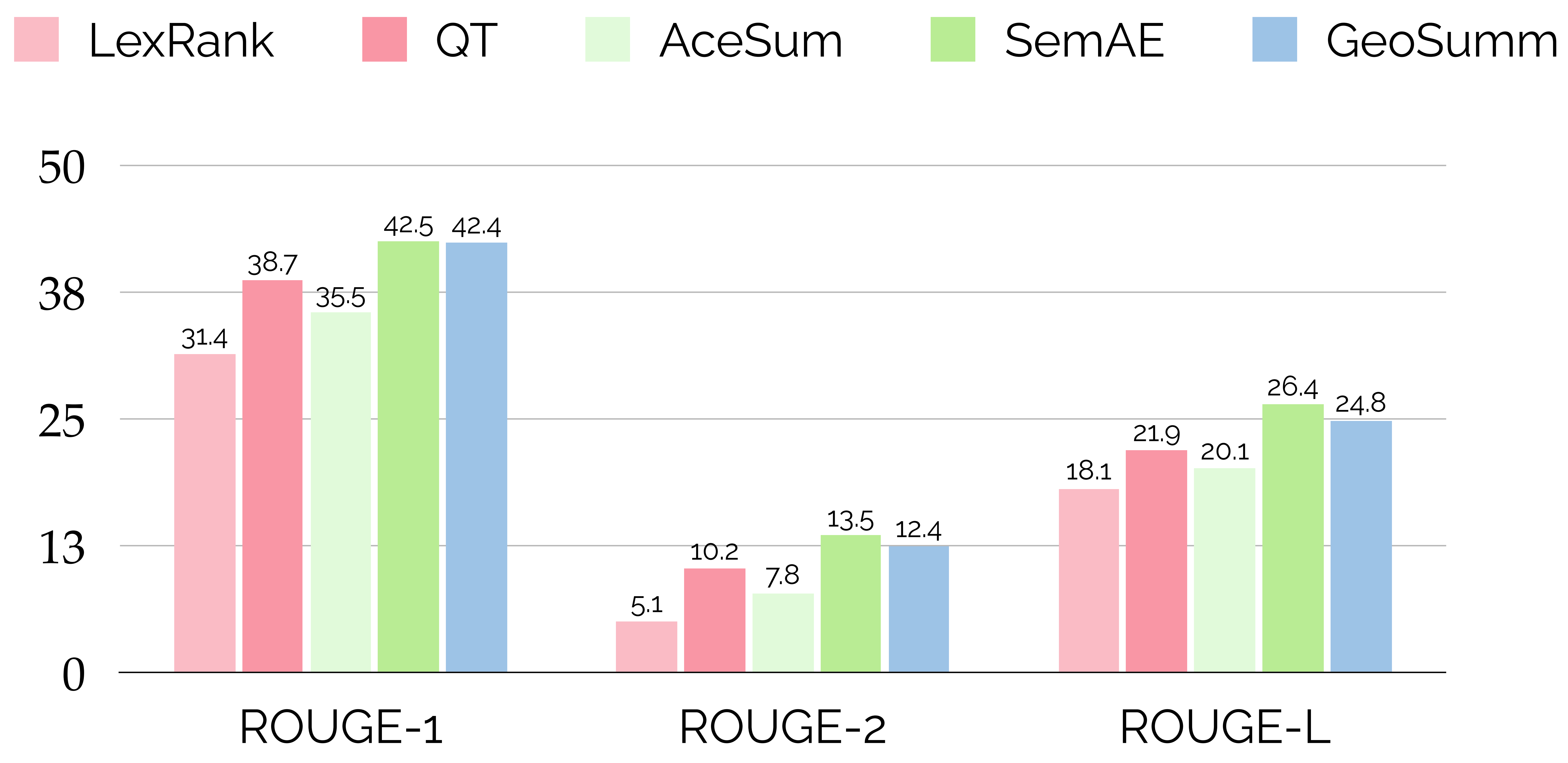
General Summarization - OPOSUM+



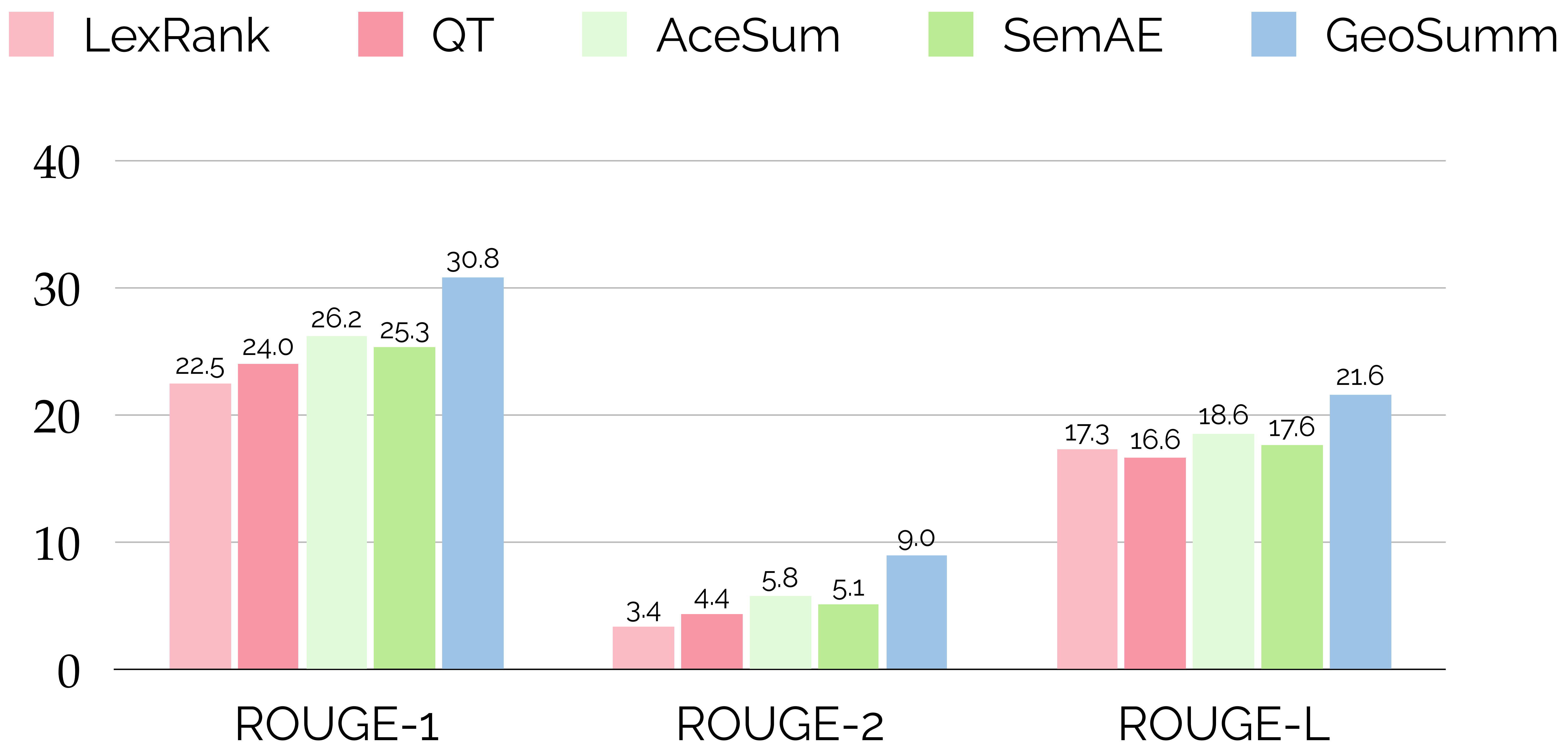
General Summarization - Amazon



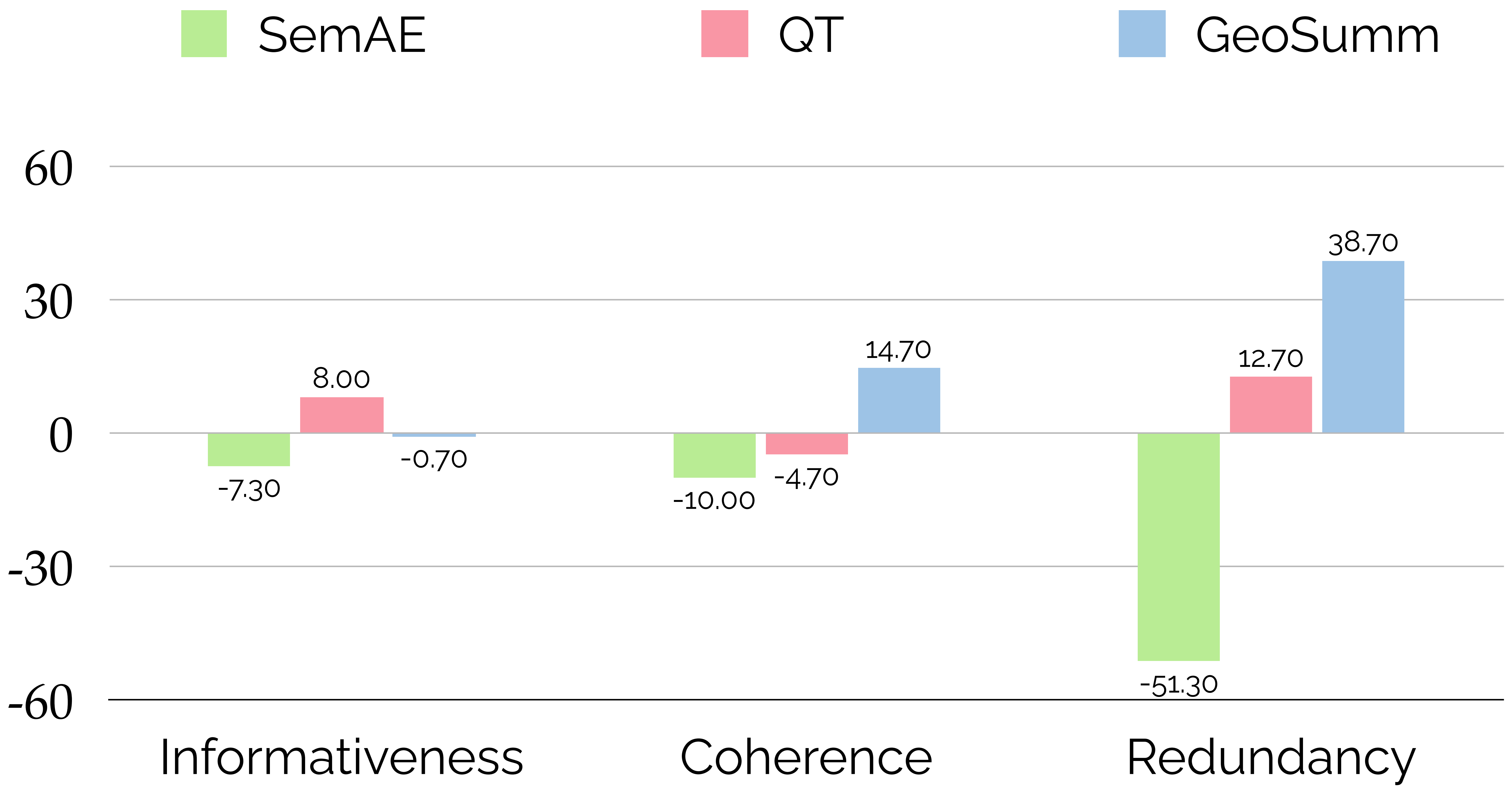
General Summarization - SPACE



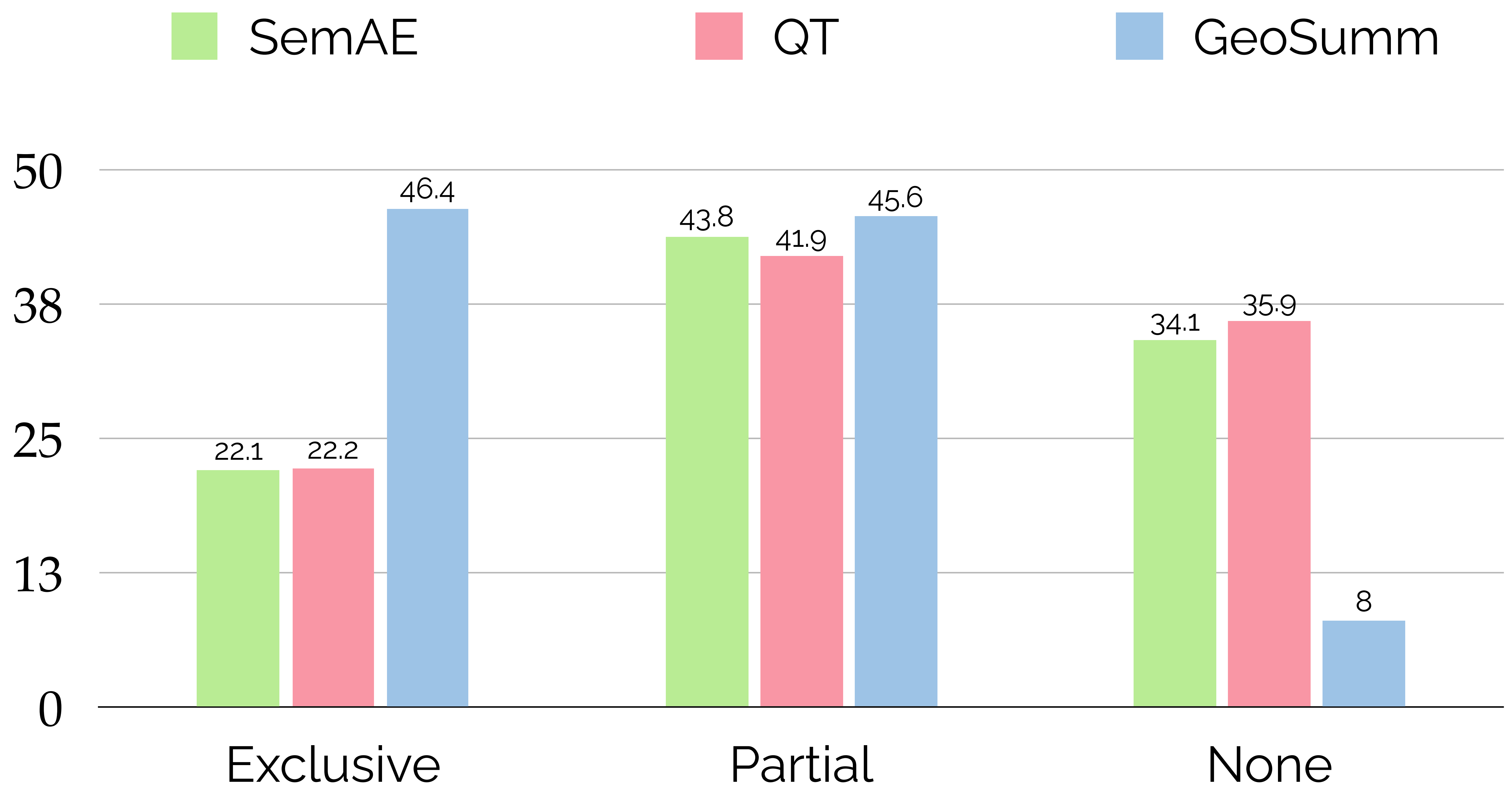
Aspect Summarization - OpoSum+



Human Evaluation - General Summaries



Human Evaluation — Aspect Summaries



Probing Representations

Cluster

- The gardens are lovely with wide variety of **flowering plants** and shrubs, koi ponds, etc.
- **Pots of tulips and daffodils** in full bloom

Flowers

Probing Representations

Cluster ●

- The gardens are lovely with wide variety of **flowering plants** and shrubs, koi ponds, etc.
- **Pots of tulips and daffodils** in full bloom

Flowers

Cluster ●

- **Calistoga** is a beautiful historic town …
- The Roman Spa and **Calistoga** is our favorite spot…

'Calistoga'

Probing Representations

Cluster

- The gardens are lovely with wide variety of **flowering plants** and shrubs, koi ponds, etc.
- **Pots of tulips and daffodils** in full bloom

Flowers

Cluster

- **Calistoga** is a beautiful historic town …
- The Roman Spa and **Calistoga** is our favorite spot…

'Calistoga'

Cluster

- The rooms were in great shape, very clean, comfortable **beds with lots of pillows**
- **The pillows and bed** coverings were of very high quality

Beds & Pillows

Output Summaries

Human	GeoSumm	SemAE	QT
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.	Overall we had a nice stay at the hotel. Our room was very clean and comfortable. The atmosphere is stylish and the service was great. We ate breakfast at the hotel and it was great. I appreciate the location and the security in the hotel. The food and service at the restaurant was awesome. The Hotel is classy and has a rooftop bar. The restaurant is cozy but they have good healthy food. Great hotel.	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.	Great hotel. We liked our room with an ocean view. The staff were friendly and helpful. There was no balcony. The location is perfect. Our room was very quiet. I would definitely stay here again. You're one block from the beach. So it must be good! Filthy hallways. Unvacuumed room. Pricy, but well worth it.

Conclusion

- GeoSumm learns topical representations from pre-trained text representations

Conclusion

- GeoSumm learns topical representations from pre-trained text representations
- GeoSumm uses them to capture salience using approximate geodesics

Conclusion

- GeoSumm learns topical representations from pre-trained text representations
- GeoSumm uses them to capture salience using approximate geodesics
- Topical representations work great, but are there better approaches?

Conclusion

- GeoSumm learns topical representations from pre-trained text representations
- GeoSumm uses them to capture salience using approximate geodesics
- Topical representations work great, but are there better approaches?
- Representations capturing varying semantics occupy different high-dimensional space

somnath@cs.unc.edu