Unsupervised Opinion Summarization Using Approximate Geodesics

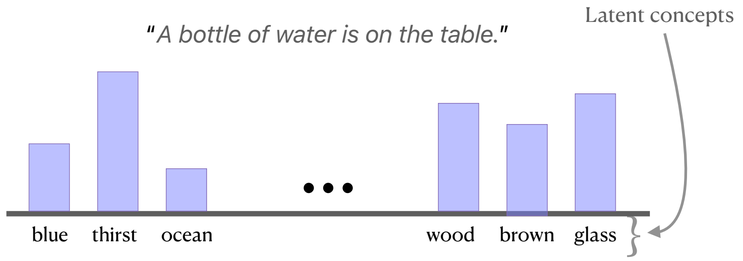
Somnath Basu Roy Chowdhury, Nicholas Monath, Avinava Dubey, Amr Ahmed, and Snigdha Chaturvedi
{somnath, snigdha}@cs.unc.edu, {nmonath, avinavadubey, amra}@google.com

Introduction

- Unsupervised opinion summarization systems are desirable due to a scarcity of labeled data
- In this work, we focus on the extractive opinion summarization task
- We propose GeoSumm that learns topical text representations from pre-trained models
- Using these topical representations, we compute saliency score of user review sentences
- We form the extractive summary based on the saliency scores

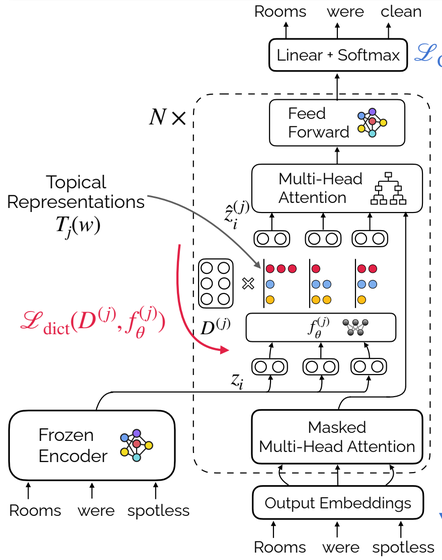
Topical Text Representations

- Topical representations capture the semantics of a text as a distribution over latent concepts.



- Being distribution over the same support, these representations can be compared using standard metrics like cosine similiarity
- Topical representations are compositional in nature allowing us to combine multiple text representations by averaging
- These properties of topical representations are useful for opinion summarization as they allow us to compute opinion saliency
- Similar operations are unavailable using distributed representation, which are often anisotropic in nature (Timkey et al, 2021) with a few dimensions dominating similarity functions.

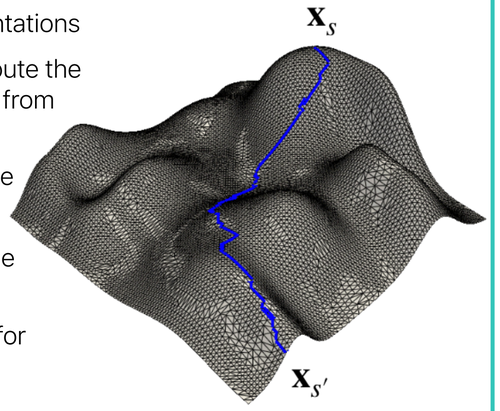
Geodesic Summarizer (GeoSumm)



- GeoSumm uses dictionary learning to decompose pre-trained text representations into topical ones.
- The dictionary captures the latent semantic units.
- The Transformer is trained using a combination of the dictionary loss and sentence reconstruction loss.
- A separate dictionary is used at every layer of the decoder.
- Sentence representation is a combination of word representations from every layer.

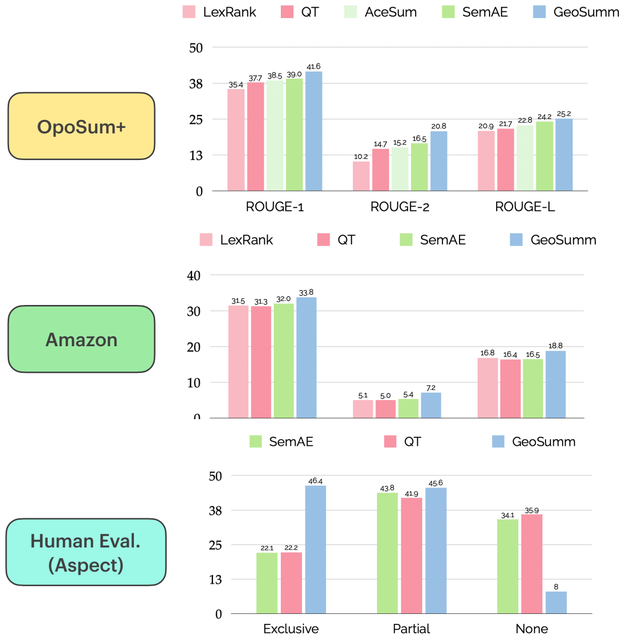
Extractive Summarization Routine

- Compute knn-graph of representations
- Use Dijkstra's algorithm to compute the geodesic distance of all reviews from the mean representation
- Use the geodesic distance as the saliency scores for each review
- Select the top k sentences as the extractive summary
- A similar approach is leveraged for aspect-based summarization



Experiments

- We perform experiments on 3 opinion summarization datasets on general and aspect-specific summarization.



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?

Checkout our paper!