

# MeanSum: A Neural Model for Unsupervised Multi-Document Abstractive Summarization

Eric Chu\*  
MIT Media Lab  
echu@mit.edu

Peter J. Liu\*  
Google Brain  
peterjliu@google.com

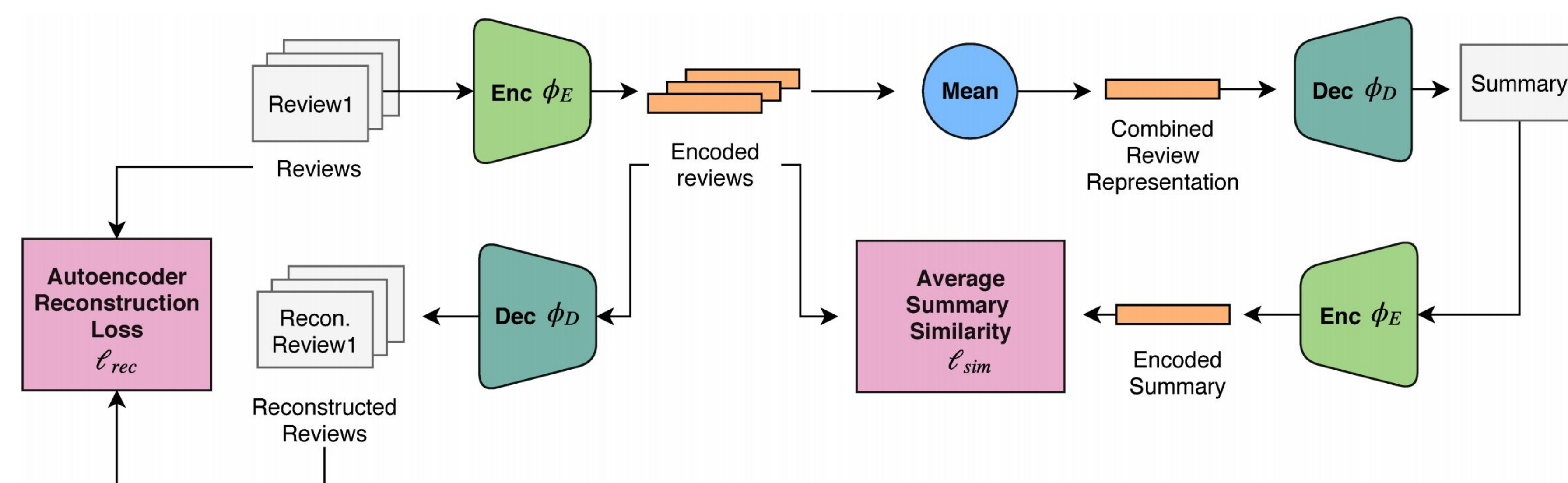
## Motivation and Task

- Majority of summarization research has relied on large, paired document-summary datasets. These datasets are rare and often do not generalize to different types of documents.
- Extractive summarization works without supervision but is limited and also not how humans summarize.
- We introduce the first end-to-end, neural, abstractive, unsupervised summarization model. Our model does not assume the existence of example summaries.
- Our model is applied to publicly available Yelp business and Amazon product reviews and summarizes multiple reviews in the form of one canonical review.
- Automated metrics — word overlap score, sentiment accuracy, and negative log likelihood — are used to guide model selection.
- We also collect a small dataset of reference summaries to use as ground truth evaluation.

## Method

### MeanSum model

- Autoencoder module** to learn review representations and constrain summaries to language domain
- Summarization module** to generate summaries that are semantically similar to input documents
  - Decoding during training is done with Straight-Through **Gumbel-Softmax estimator** (Jang et al, 2017)
- Encoders and decoders are tied (\*decoders must be tied to work\*)
  - Initialized with pre-trained language model



### Automated metrics without summaries

- Word Overlap:** is the summary on-topic?
  - Average ROUGE-1 between summary and reviews:  $\frac{1}{N} \sum_{i=1}^N \left[ \frac{1}{k} \sum_{j=1}^k \text{ROUGE}(s^{(i)}, x_j^{(i)}) \right]$
- Sentiment accuracy:** does the summary reflect overall sentiment of original reviews?
  - Pre-train a 1-5 star rating classifier.
  - Across data points, average the binary accuracy of whether the summary's predicted rating is equal to the mean rating of reviews:  $\frac{1}{N} \sum_{i=1}^N \left[ \text{CLF}(s^{(i)}) = \text{round} \left( \frac{1}{k} \sum_{j=1}^k r(x_j^{(i)}) \right) \right]$
- Negative log-likelihood (NLL):** is the summary fluent?
  - Measured using pre-trained language model

## Conclusion

- We show neural abstractive summarization is possible without examples.
- Model is competitive or better than recent, near SOTA extractive model according to automated metrics, human evaluation, and metrics on ground-truth reference summaries.
- Open-source: models and code task implemented in <https://github.com/sosuperic/MeanSum>

## Results and Examples

		Vs. Reference Summaries			Metrics Without Summaries			
		Model	ROUGE-1	ROUGE-2	ROUGE-L	Sentiment Acc.	Word Overlap	NLL
Baselines	MeanSum (ours)		28.86	3.66	15.91	<b>51.75</b>	26.09	1.19
	Extractive		24.61	2.85	13.81	42.95	28.59	--
	No training		21.22	1.69	11.92	24.44	19.68	1.29
	Best review		27.97	3.46	15.29	38.48	23.86	--
	Worst review		16.91	1.66	11.11	30.01	13.14	--
Model Variants	Multi-Lead-1		26.79	<b>3.77</b>	14.39	40.69	31.64	--
	No pre-trained LM		26.16	3.07	15.31	48.97	23.67	<b>1.14</b>
	No autoencoder		--	--	--	--	--	--
	Rec cycle loss		25.23	3.58	15.82	43.65	22.26	<b>1.14</b>
	Early cosine loss		14.35	1.26	9.02	19.32	14.28	1.71
	Untied decoders		--	--	--	--	--	--
	Untied encoders		<b>29.35</b>	3.52	<b>15.97</b>	50.89	26.29	1.20

### Takeaways:

- MeanSum outperforms extractive model on ROUGE using reference summaries
- MeanSum and extractive summaries rated comparably in terms of sentiment, information, and fluency scores
- Average fluency rated close to human reviews
- Ablation study shows that no autoencoder and untied decoders variants result in degenerate summaries
- Model is robust to varying number of reviews  $k$  at train and test time

### Mechanical Turk results evaluating quality

Model	MeanSum (ours)	Extractive	Random review
Sentiment	<b>3.91</b>	3.87	--
Information	3.83	<b>3.85</b>	--
Grammar	<b>3.97</b>	3.86	3.94
Non-redundancy	3.74	3.93	<b>4.06</b>
Referential clarity	<b>4.13</b>	4.05	4.09
Focus	4.10	4.01	<b>4.23</b>
Structure	<b>4.02</b>	3.99	4.01

### Sample Model Input/Output

#### Original Reviews: Mean Rating = 4

- No question the **best pedicure** in Las Vegas. I go around the world to places like Thailand and Vietnam to get beauty services and this place is the real thing. Ben, Nancy and Jackie took the time to do it right and **you don't feel rushed**. My cracked heels have never been softer thanks to Nancy and they didn't hurt the next day.
- Came to Vegas to visit sister both wanted full sets got to the salon like around 4. **Friendly** guy greet us and ask what we wanted for today but girl doing nails was very rude and immediately refuse service saying she didn't have any time to do 2 full sets when it clearly said open until 7pm!
- This is the most clean nail studio I have been so far. The service is great. **They take their time** and do the **irk with love**. That creates a very **comfortable atmosphere**. I recommend it to everyone!!
- Took a taxi here from hotel bc of reviews -Walked in and walked out - not sure how they got these reviews. Strong smell and broken floor - below standards for a beauty care facility.
- The **best** place for pedi in Vegas for sure. My husband and me moved here a few months ago and we have tried a few places, but this is the only place that makes us 100% happy with the result. I highly recommend it!
- This was the **best** nail experience that I had in awhile. The service was perfect from start to finish! I came to Vegas and needed my nails, feet, eyebrows and lashes done before going out. In order to get me out quickly, my feet and hands were done at the same time. Everything about this place was excellent! I will certainly keep them in mind on my next trip.
- I came here for a much needed **pedicure** for me and my husband. We got **great customer service** and an amazing **pedicure and manicure**. I will be back every time I come to Vegas. My nails are beautiful, my skin is very soft and smooth, and most important I felt great after leaving!!!
- My friend brought me here to get my very first **manicure** for my birthday. Ben and Nancy were so **friendly** and super attentive. Even though were there past closing time, **I never felt like we were being rushed or that they were trying to get us out the door**. I got the #428 Rosewood gel **manicure** and I love it. I'll definitely be back and next time I'll try a **pedicure**.

#### Abstractive Summary: Predicted Rating = 5

Probably the **best mani/pedi** I have ever had. I went on a Saturday afternoon and it was busy and they have a great selection of colors. We went to the salon for a few hours of work, but this place was **very relaxing**. **Very friendly staff** and a great place to relax after a long day of work.

#### Extractive Summary: Predicted Rating = 1

Came to Vegas to visit sister both wanted full sets got to the salon like around 4. **Friendly** guy greet us and ask what we wanted for today but girl doing nails was very rude and immediately refuse service saying she didn't have any time to do 2 full sets when it clearly said open until 7pm!