

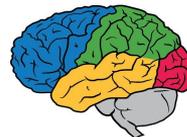
MeanSum: A Neural Model for Unsupervised Multi-Document Abstractive Summarization

Eric Chu*

MIT Media Lab

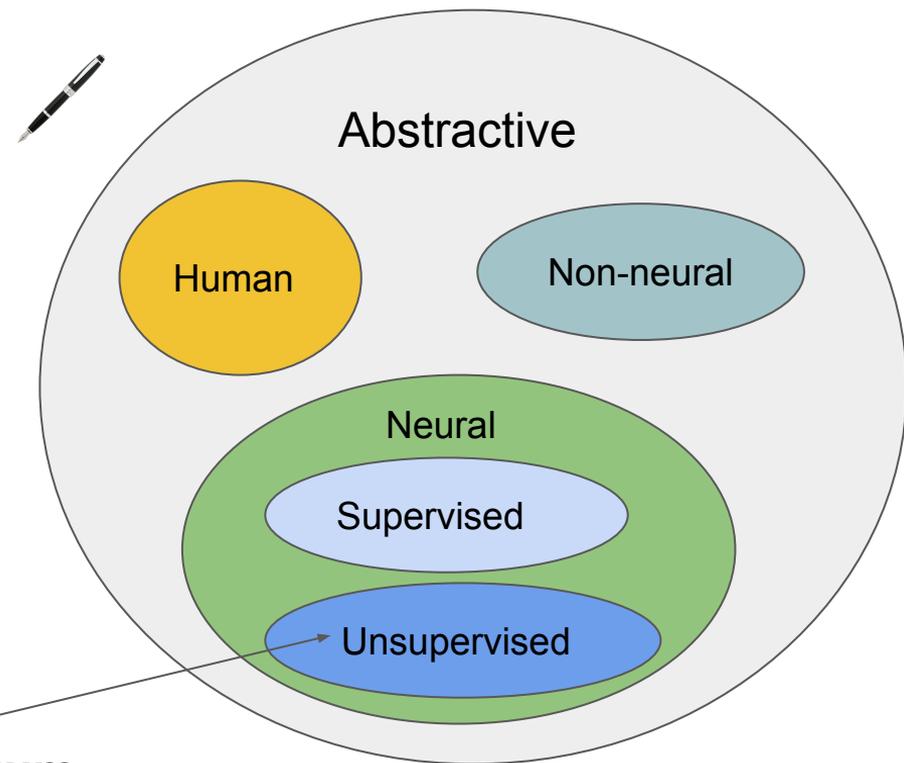
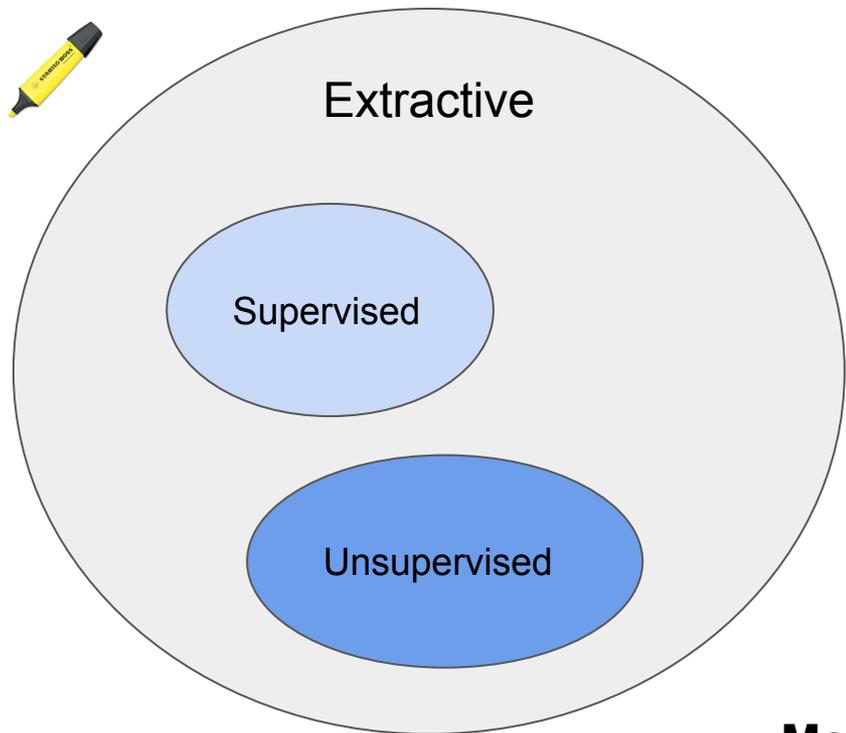
Peter J. Liu*

Google Brain



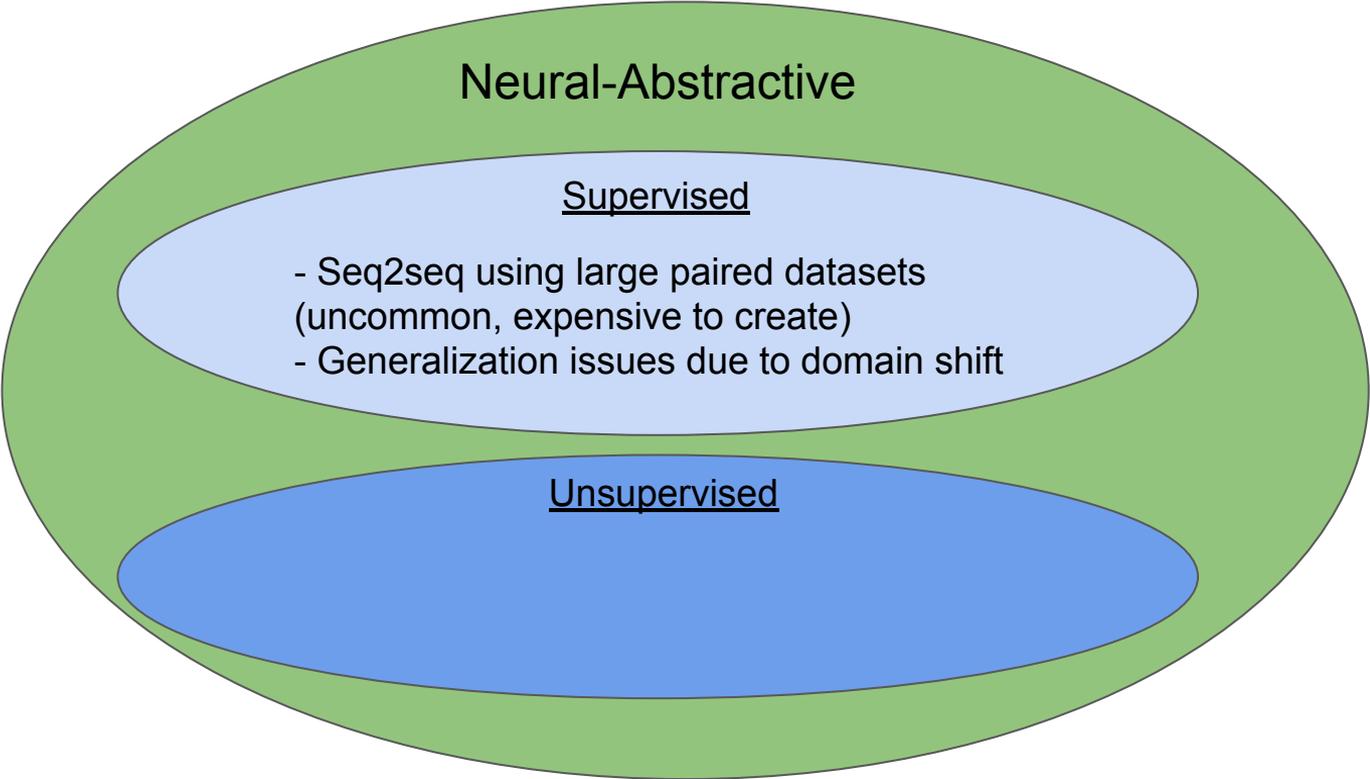
ICML 2019

Text summarization overview



MeanSum

Neural-Abstractive

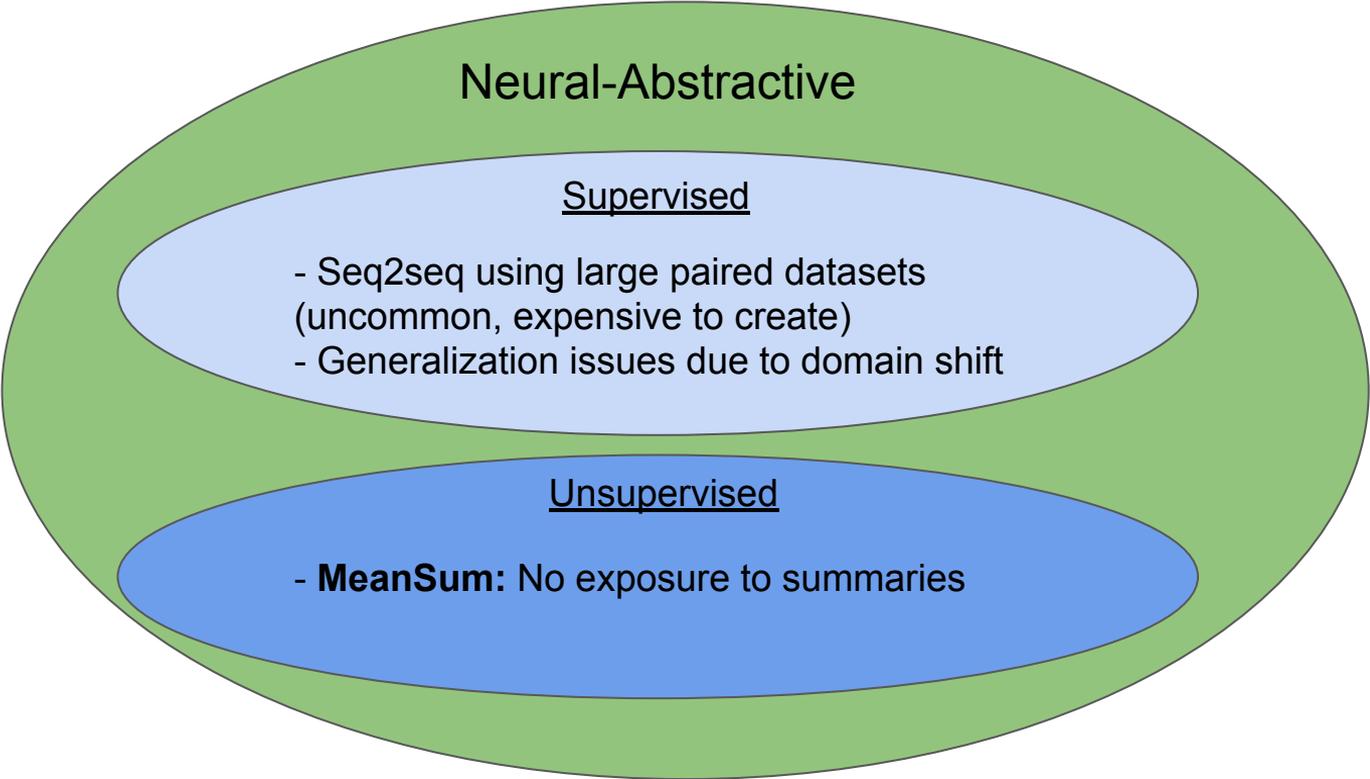


Supervised

- Seq2seq using large paired datasets (uncommon, expensive to create)
- Generalization issues due to domain shift

Unsupervised

Neural-Abstractive



Supervised

- Seq2seq using large paired datasets (uncommon, expensive to create)
- Generalization issues due to domain shift

Unsupervised

- **MeanSum**: No exposure to summaries

Dataset and task

Summarize reviews for a product or business (Yelp/Amazon)

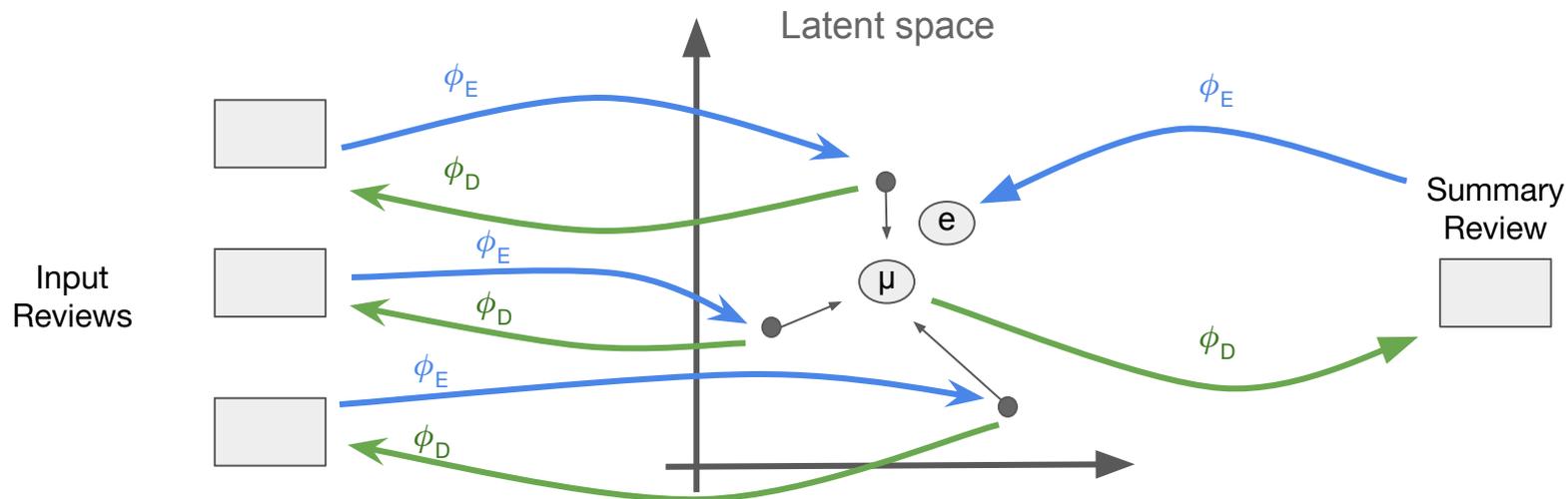
Example Input (nail salon):

“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.”

“This is the most clean nail studio I have been so far. The service is great. They take their time and do the work with love. That creates a very comfortable atmosphere. I recommend it to everyone!!”

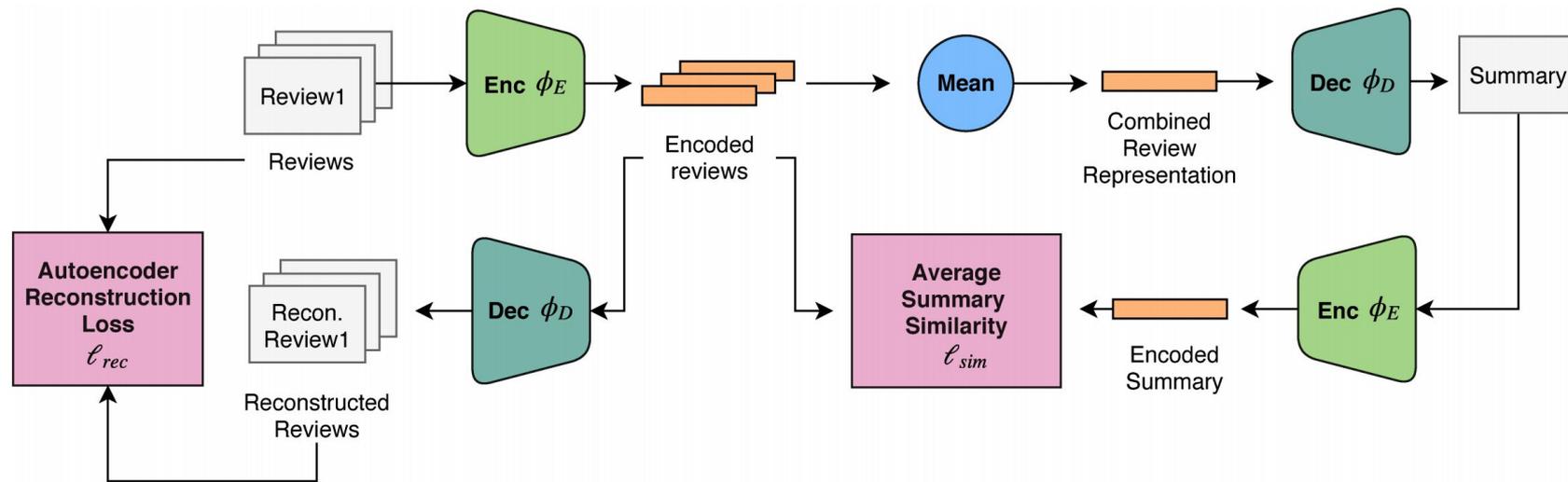
“The best place for pedi in Vegas for sure. My husband and me moved here a few months ago and we have tried a few places, but this is the only place that makes us 100% happy with the result. I highly recommend it!”

High-level idea

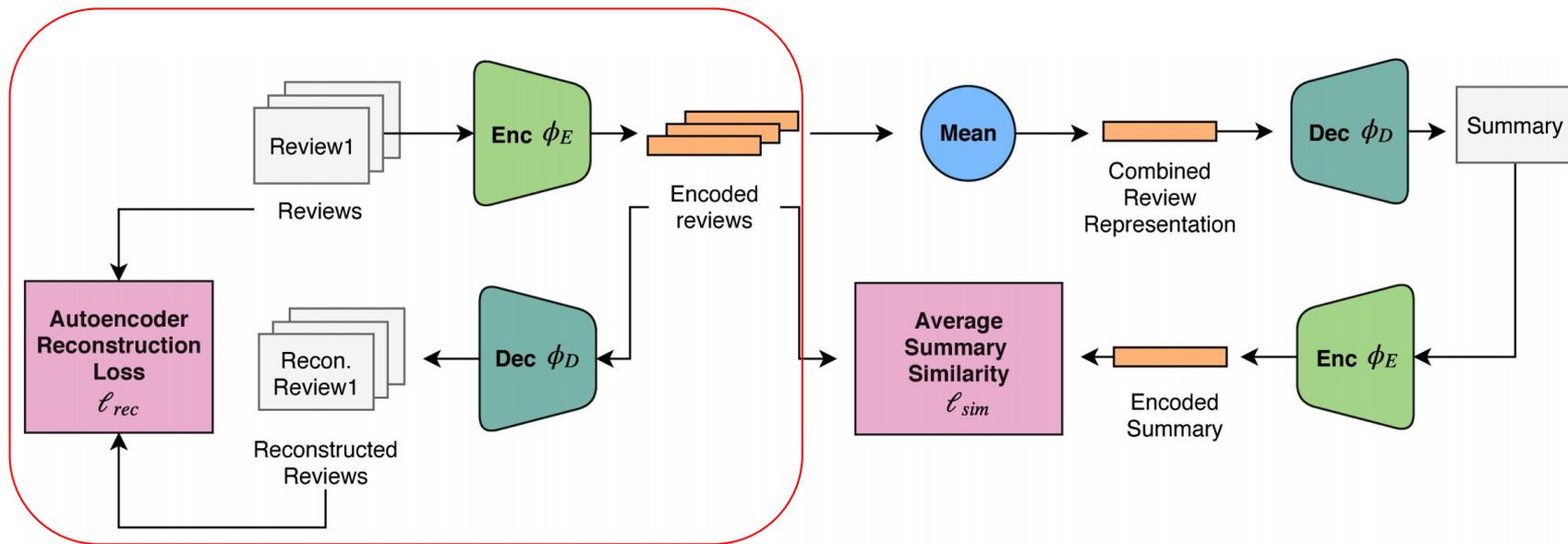


Autoencoder: (encoder, decoder) = (ϕ_E, ϕ_D)

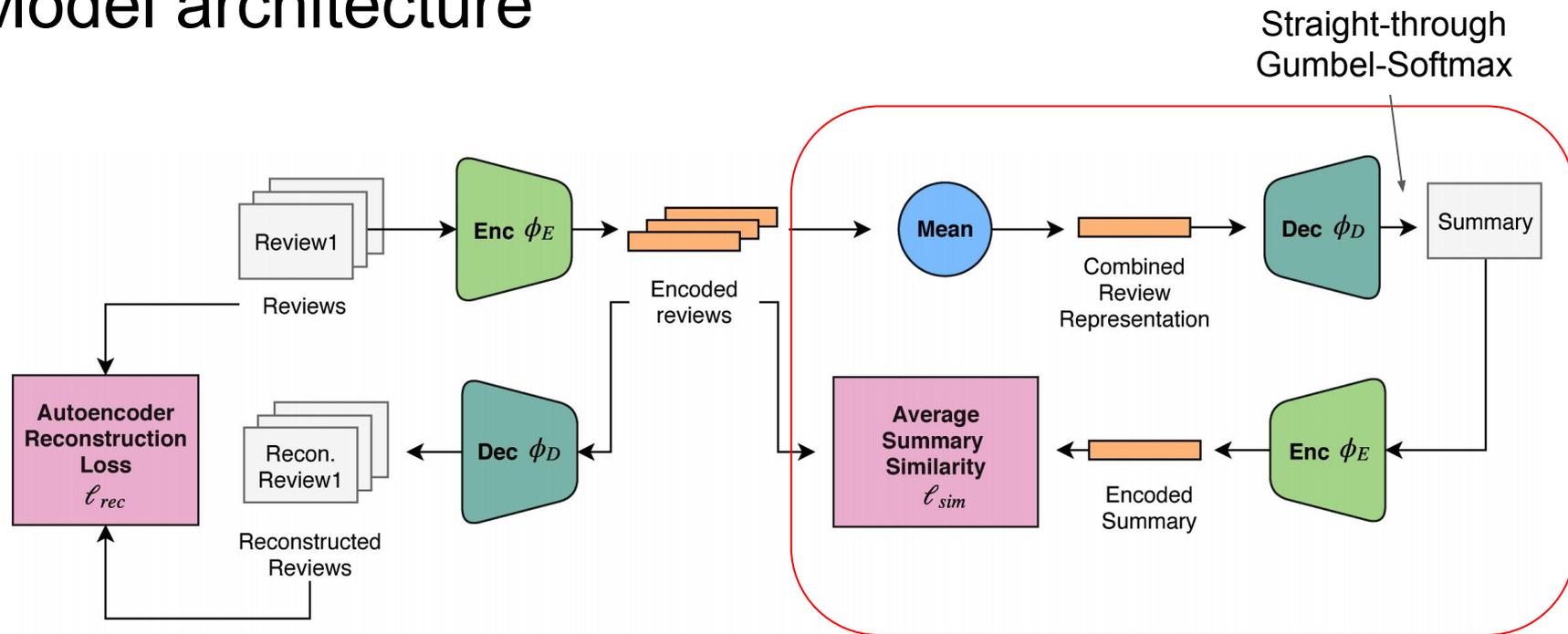
Model architecture



Model architecture



Model architecture



$$\ell_{model} = \ell_{rec} + \ell_{sim}$$

Proxy metrics for tuning models without ground-truth

1. Sentiment accuracy, using pretrained rating classifier
2. Average word overlap with input reviews (relevance)
3. Negative Log-likelihood (NLL), using pretrained language model (fluency)

Results (automatic metrics)

Model		<i>Vs. Reference Summaries</i>			<i>Metrics Without Summaries</i>		
		ROUGE-1	ROUGE-2	ROUGE-L	Sentiment Acc.	WO	NLL
MeanSum (ours)		28.86	3.66	15.91	51.75	26.09	1.19
<i>Baselines</i>	Extractive (Rossiello et al., 2017)	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	–
	Multi-Lead-1	26.79	3.77	14.39	40.69	31.64	–
<i>Model Variants</i>	No pre-trained language model	26.16	3.07	15.31	48.97	23.67	1.14
	No auto-encoder	–	–	–	–	–	–
	Reconstruction cycle loss	25.23	3.58	15.82	43.65	22.26	1.14
	Early cosine loss	14.35	1.26	9.02	19.32	14.28	1.71
	Untied decoders	–	–	–	–	–	–
	Untied encoders	29.35	3.52	15.97	50.89	26.29	1.20

Table 1. Automated metric results with $k = 8$ reviews being summarized. The reference summaries results are shown for the test split. Note for Best/Worst Review WO scores: we exclude the best/worst review when calculating the average. Numbers are not provided for models that degenerated into non-natural language. As noted earlier, the NLL’s are only provided for our abstractive models.

Results (human evaluation)

	Model	Sentiment	Information		
	MeanSum (ours)	3.91	3.83		
	Extractive	3.87	3.85		
Model	Grammar	Non- redundancy	Referential clarity	Focus	Structure and Coherence
MeanSum (ours)	3.97	3.74	4.13	4.10	4.02
Extractive	3.86	3.93	4.05	4.01	3.99
Early cosine loss	2.02	1.84	2.02	1.96	1.95
Random review	3.94	4.06	4.09	4.23	4.01

Table 2. Mechanical Turk results evaluating quality of summaries.

Original Reviews: Mean Rating = 4

1: 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.

2: 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!

3: 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!!

4: 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.

5: 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!

6: 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.

7: 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!!!

8: 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 we 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**.

MeanSum Model: 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 Model: 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!

Conclusion

- We showed neural abstractive summarization is possible without examples!
- Limitations:
 - Summary has the shape/form of an input document
 - Multi-document model
- Models and code available online:

<https://github.com/sosuperic/MeanSum>

Come talk to us at our poster: Thursday June 13, 06:30 -- 09:00 PM, Pacific Ballroom