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

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
  - Decoders and decoders are tied (“decoders must be tied to work”)
  - Initialized with pre-trained language model

Results and Examples

- Vs. Reference Summaries

<table>
<thead>
<tr>
<th>Model</th>
<th>ROUGE-1</th>
<th>ROUGE-2</th>
<th>ROUGE-L</th>
<th>Sentence Acc.</th>
<th>Word Overlap</th>
<th>NLL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeanSum (ours)</td>
<td>28.86</td>
<td>3.66</td>
<td>15.91</td>
<td>51.75</td>
<td>26.09</td>
<td>1.19</td>
</tr>
<tr>
<td>Extractive</td>
<td>24.61</td>
<td>2.85</td>
<td>13.81</td>
<td>42.95</td>
<td>28.59</td>
<td>--</td>
</tr>
<tr>
<td>No training</td>
<td>21.22</td>
<td>1.69</td>
<td>11.92</td>
<td>24.44</td>
<td>19.66</td>
<td>1.29</td>
</tr>
<tr>
<td>Best review</td>
<td>27.97</td>
<td>3.46</td>
<td>15.29</td>
<td>38.48</td>
<td>23.86</td>
<td>--</td>
</tr>
<tr>
<td>Worst review</td>
<td>16.91</td>
<td>1.66</td>
<td>11.11</td>
<td>30.01</td>
<td>13.14</td>
<td>--</td>
</tr>
<tr>
<td>Multi-Lead-1</td>
<td>26.79</td>
<td>3.77</td>
<td>14.39</td>
<td>40.69</td>
<td>31.94</td>
<td>--</td>
</tr>
<tr>
<td>No pre-trained LM</td>
<td>26.18</td>
<td>3.07</td>
<td>15.31</td>
<td>48.97</td>
<td>23.67</td>
<td>1.14</td>
</tr>
<tr>
<td>No autoencoder</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Rec cycle loss</td>
<td>25.23</td>
<td>3.58</td>
<td>15.82</td>
<td>43.65</td>
<td>22.26</td>
<td>1.14</td>
</tr>
<tr>
<td>Early cosine loss</td>
<td>14.35</td>
<td>1.26</td>
<td>9.02</td>
<td>19.32</td>
<td>14.28</td>
<td>1.71</td>
</tr>
<tr>
<td>Untied decoders</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Untied encoders</td>
<td>29.35</td>
<td>3.52</td>
<td>15.97</td>
<td>50.89</td>
<td>26.29</td>
<td>1.20</td>
</tr>
</tbody>
</table>

- Metrics Without Summaries

<table>
<thead>
<tr>
<th>Model</th>
<th>MeanSum (ours)</th>
<th>Extractive</th>
<th>Random review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentiment</td>
<td>3.91</td>
<td>3.87</td>
<td>--</td>
</tr>
<tr>
<td>Information</td>
<td>3.83</td>
<td>3.85</td>
<td>--</td>
</tr>
<tr>
<td>Grammar</td>
<td>3.97</td>
<td>3.86</td>
<td>3.94</td>
</tr>
<tr>
<td>Non-redundancy</td>
<td>3.74</td>
<td>3.93</td>
<td>4.08</td>
</tr>
<tr>
<td>Referential clarity</td>
<td>4.13</td>
<td>4.05</td>
<td>4.09</td>
</tr>
<tr>
<td>Focus</td>
<td>4.10</td>
<td>4.01</td>
<td>4.23</td>
</tr>
<tr>
<td>Structure</td>
<td>4.92</td>
<td>3.99</td>
<td>4.01</td>
</tr>
</tbody>
</table>

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

- MeanSum outperforms extractive model on ROUGE using reference summaries
- MeanSum and abstractive 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

Takeaways:

- MeanSum outperforms extractive model on ROUGE using reference summaries
- MeanSum and abstractive 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

Abstractive Summary: Predicted Rating = 5

Probably the best manicure!! 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.

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. I got the #428 Rosewood gel manicure and it was amazing. I go around the world to places like Thailand and Vietnam to get beauty services and this place was the best manicure that I had in awhile. The service was perfect from start to finish. I come 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 done at the same time. Everything about this place was excellent! I will certainly keep them in mind on my next trip.
3. This was the most pleasant and relaxing experience that I have had. The service was great and the ambiance was perfect. I highly recommend it to everyone!
4. This is the most clean nail studio I have been so far. The service is great. I recommend it to everyone!!
5. This is the best salon I have been to in the entire world. I go around the world to places like Thailand and Vietnam to get beauty services and this is the only place that makes us 100% happy with the result. I highly recommend it. I highly recommend this place.

Mechanical Turk results evaluating quality

- Sample Model Input/Output

- 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. I got the #428 Rosewood gel manicure and it was amazing. I go around the world to places like Thailand and Vietnam to get beauty services and this place was the best manicure that I had in awhile. The service was perfect from start to finish. I come 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 done at the same time. Everything about this place was excellent! I will certainly keep them in mind on my next trip.
3. This was the most pleasant and relaxing experience that I have had. The service was great and the ambiance was perfect. I highly recommend it to everyone!
4. This is the most clean nail studio I have been so far. The service is great. I recommend it to everyone!!
5. This is the best salon I have been to in the entire world. I go around the world to places like Thailand and Vietnam to get beauty services and this is the only place that makes us 100% happy with the result. I highly recommend it. I highly recommend this place.

Sample Model Input/Output

- Abstractive Summary: Predicted Rating = 5

Probably the best manicure!! 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.

Abstractive Summary: Predicted Rating = 5

Probably the best manicure!! 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.