
Unsupervised summarization of psychotherapy data

Aaron Z. Reed
Department of Physics
Stanford University
aaron73@stanford.edu

Abstract

A shortage of psychotherapists in the US and the disruption in therapy practice caused by the COVID-19 pandemic motivate the development of high-quality therapy chatbots. Chatbots that produce a summary of their conversations with clients would provide valuable data to therapists and enable more effective treatment. Here we compare two methods of unsupervised summarization of therapy session transcripts: an extractive approach and an abstractive approach. We analyze the results of each approach both subjectively and with automated metrics, and conclude that both extractive and abstractive techniques are promising approaches to automated summarization in the context of psychotherapy.

1 Introduction

There are projections of a shortage of therapists in the US, and university CAPS departments all around the US are often overwhelmed and underfunded, leaving many students waiting for long periods of time before they are able to attend their appointments with therapists. [1] Coinciding with this is the rise of mental health chatbots aimed to provide therapeutic assistance to users, such as Woebot [2] and Wysa. [3] The use of deep-learning methods to summarize high-quality therapy data would help integrate these mental health chatbots into more robust psychotherapy programs as supplements; chatbots that engage with users in between real therapy sessions can provide therapists with summaries of the content of said conversations, which can help reduce therapists' workload and provide clients with better quality of care.

Motivated by this problem, our original goal was to design an algorithm that would summarize the transcript of an approximately hour-long psychotherapy session into a few sentences or a paragraph for the therapist's future reference. We initially intended to use a corpus of psychotherapy transcripts that included labeled summaries in the context of a supervised approach to summarization. [4]

However, we were unable to obtain access to this dataset, so we resorted to using a publicly available dataset compiled from Counsel Chat, a website where users can ask questions that are answered by qualified therapists. [5] [6] The dataset contains around 2,000 questions with their responses provided as a CSV file, but there are no labels or summaries included with the dataset. Therefore, we had to reframe our goal in terms of *unsupervised* text summarization.

We now define our learning task as follows: we take as input a question asked by a user and a group of responses to it by qualified therapists. Using unsupervised learning methods (detailed in the next section), we generate a summary of the responses in the form of an answer that tries to capture a general consensus among the responses.

2 Related work

Unsupervised text summarization can be broadly divided into *extractive* and *abstractive* approaches. [7] [8] Extractive techniques draw vocabulary exclusively from the text to be summarized, whereas abstractive techniques use larger vocabularies and can include language in the summary that is not found in the original text.

2.1 Extractive summarization

A typical deep learning method for extractive summarization is word embedding, which maps a vocabulary of V words to a vector space of dimension $N < V$. The vector in the N -dimensional space representing a particular word is called that word's *embedding*. Ideally, the relationships between embeddings should reflect the meaning of the words they encode, which can be learned by using the context of the words as a proxy for their meaning.¹ We examine a context window of n surrounding words and map words with similar contexts to similar embeddings. We can then train a recurrent neural net, referred to as the *encoder*, using the Word2Vec algorithm. [9] Roughly speaking, Word2Vec asks the encoder to predict a target word given its context and then compares the prediction to the actual missing word in order to train the encoder. The context with the target word missing is called a *skip-gram*, thus training an encoder on skip-grams is called *skip-gram encoding*, and the resulting word embeddings are called *skip-gram vectors*.

One variant of Word2Vec for summarization is Sent2Vec, where entire sentences are encoded into so-called *skip-thought* vectors and context consists of the n preceding and following *sentences* in a text. This approach is taken by [10], who use skip-thoughts to train an encoder-decoder model that maps sentences to vectors and then predicts surrounding sentences. The SkipThoughts model forms the basis of our baseline extractive summarization model, explained in Section 4. The architecture of the SkipThoughts model is shown in Figure 1.

Another extractive algorithm that takes a similar approach is called QuickThoughts. [12] The authors built on the work of [10] by elaborating the encoder/decoder architecture to achieve an order-of-magnitude speedup in training time. We intend to compare the results as well as the training times of QuickThoughts to the baseline of SkipThoughts from which it derives.

2.2 Abstractive summarization

Abstractive architectures are generally more complex than extractive architectures. This is necessary in order to give the abstractive model the ability to draw from a wider vocabulary beyond the specific example it is summarizing.

A case-in-point of this increased complexity is the MeanSum algorithm. [13] As shown in Figure 2, the model includes an auto-encoder on the left and a summarizer on the right. The weights of the encoders and decoders are tied (i.e., set equal to each other), so the auto-encoder exerts a normalizing effect on the summary generator by ensuring that the sentence embeddings decode to something representative of the original review. On the summary generator side, the embeddings of many texts (here, Yelp reviews) are averaged together and the average embedding vector is then decoded into a summary text.

Notably, the summary can contain vocabulary from ANY of the corpus used to train the encoder/decoder, not only the specific example. This ability to “go beyond” the vocabulary of the specific example constitutes the *abstractive* nature of the MeanSum algorithm.

3 Dataset and Features

As mentioned above, our dataset consists of publically available data scraped from the website Counsel Chat. [5] The data is a CSV where each row is a single answer, along with the name of the therapist answering and the question being answered. A portion of the CSV is shown in Figure 3.

We preprocessed our dataset differently depending on the algorithm to which it was fed. For SkipThoughts, the original input was a list containing the text of each text to be summarized

¹Put another way, we are use *syntactic* agreement as a proxy for *semantic* agreement.

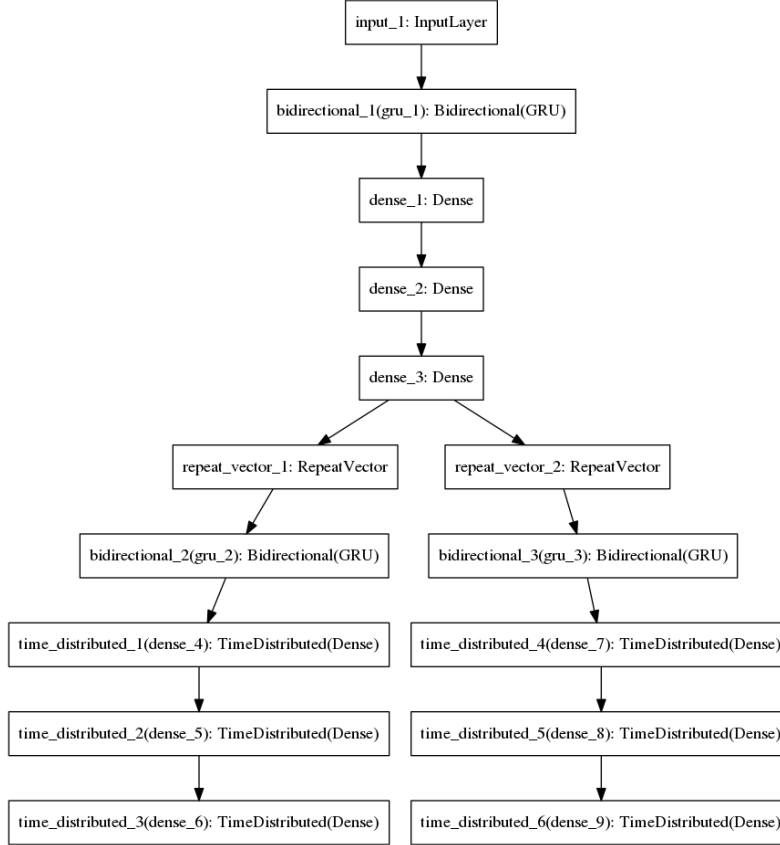


Figure 1: Model architecture of the pre-trained skip-thought network, as presented in [11]

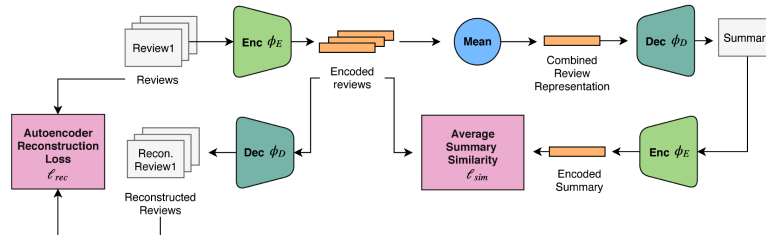


Figure 2: MeanSum model architecture with autoencoder (left half) and summary generator (right half). [13]

(originally, a list of emails). Thus, we simply unpacked the CSV, selected the column containing therapist answers, and did string processing.

In the case of MeanSum, the original input was several JSON files containing Yelp reviews and associated metadata. We would have had to create a similar data structure and insert our CounselChat data in place of the Yelp data.

4 Methods

4.1 SkipThoughts

As our baseline against which to measure the other two algorithms, we first computed summaries using the SkipThoughts algorithm. We modified a previously-published application of the SkipThoughts algorithm to email summarization [11] so that it would run on CounselChat data. Since it was

questionText	answerText
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless and how I shouldn't be here. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?	If everyone thinks you're worthless, then maybe you're worthless too. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless and how I shouldn't be here. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?	Hello, and thank you for your question and seeking help. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless and how I shouldn't be here. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?	First thing I'd suggest is getting the sleep you need. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless and how I shouldn't be here. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?	Therapy is essential for those that are feeling depressed. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless and how I shouldn't be here. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?	I first want to let you know that you are not alone in feeling this way. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless and how I shouldn't be here. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?	Heck, sure thing, hun! Feelings of 'depression' have a way of making you feel like you're worthless. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless and how I shouldn't be here. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?	You are exhibiting some specific traits of a particular mental health condition. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?
I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless and how I shouldn't be here. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?	That is intense. Depression is a liar. Sometimes depression can make you feel like you're worthless. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?

Figure 3: One question and several associated answers from the CounselChat dataset. [5]

intended as a baseline metric, we did no training and simply used weights that had been previously trained on the open-source Enron email dataset. [14]

4.2 MeanSum

We are in the process of evaluating MeanSum on CounselChat data. For some of the technical difficulties we have encountered, see Section 6.2.

5 Experiments/Results/Discussion

The first question in the CounselChat dataset reads:

I'm going through some things with my feelings and myself. I barely sleep and I do nothing but think about how I'm worthless and how I shouldn't be here. I've never tried or contemplated suicide. I've always wanted to fix my issues, but I never get around to it. How can I change my feeling of being worthless to everyone?

SkipThoughts generates the following summary (with omissions) for the answers to this question:

The new thought may feel worse than the last one! You can always change your feelings and change your way of thinking by being open to trying to change. You deserve the help of someone helping you change your feelings of worthlessness. [...] Wishing you the absolute best! [...] National Suicide Prevention Lifeline is 1-800-273-8255. [...] It sounds like you may be struggling with depression. Anything that shows yourself you are worthy of this effort. What are the Benefits and Costs? We start with noticing. Beginning to build a better relationship and understanding of yourself will deeply influence the relationships around you. (We usually do it over a few sessions.)

We notice that the phrases make sense individually and seem to flow well in the beginning of the output, but do not reach a sense of closure at the end of the summary. For example, the latter part of the summary contains what sounds like a send-off (“Wishing you the absolute best!”) which is then followed by additional information and some *non sequitur* sentences. Comparing with the results of SkipThoughts’ email summaries, which presented sentences in the original order of the email, we diagnosed the problem as one of sentence ordering in the input. We could have rectified the issue by feeding the data in the format of all 23 first sentences, followed by all 23 second sentences, etc.

For quantitative results, we intended to compare the ROUGE scores of the different models' outputs, but we only have the SkipThoughts result so far.

We had also planned to train each model (SkipThoughts, QuickThoughts, MeanSum) on CounselChat data and compare the training speed (esp. QuickThoughts vs. SkipThoughts), but ran out of time to do so.

6 Conclusion, Challenges, and Future Work

6.1 SkipThoughts

We successfully adapted the SkipThoughts model to produce summaries of CounselChat datasets, interpreted the results in terms of the email output, and hypothesized ways to improve output quality. We intended to use these results as a baseline and compare them to the output of the other algorithms using BLEU or ROUGE scoring, but were unable to do so because we could not generate the comparison summaries in time.

Due to constraints of time and team size, we were unable to train any of the models and had to resort to evaluation using the pre-trained weights included with each of them. Given more time and personnel, we could have begun from the pre-trained weights and trained them for longer on our therapy-specific datasets. This may have mitigated some of the difference in SkipThoughts performance between the Enron email dataset and the CounselChat dataset, for example.

6.2 MeanSum

Unfortunately, even after a significant investment of time (~ 1 week), we were unable to evaluate the MeanSum abstractive summarization algorithm in comparison with SkipThoughts and QuickThoughts. We first tried running MeanSum directly on our own machines, but discovered that its dependencies required rollback of numerous core Python packages (including numpy, scipy, matplotlib, and Python itself). We avoided this rollback because it would have broken much of our existing software (including parts of this project!), and instead containerized the MeanSum application using Docker². We attempted to run the containerized algorithm both locally and in a hosted runtime (Google Colab), but the data preprocessing step of the algorithm exceeded RAM limits and was killed by the OS kernel in both cases. By this point we were out of time, and thus we were unable to present a comparison of an abstractive summarization method with the two extractive methods as we had originally intended.

Despite its current failure, our progress on the MeanSum evaluation presents an opportunity for future work. Our next step is to run MeanSum inside a virtual machine in a cloud compute instance (probably AWS), which will allow us to allocate sufficient RAM to the VM to ensure that MeanSum will run.

7 Code

Code for SkipThoughts, QuickThoughts, and MeanSum experiments can be found in the following Google Colab notebooks.

SkipThoughts: https://colab.research.google.com/drive/1B1K5Jxn7BQQwQhYhd-q6nGD9_wVbIYuq?usp=sharing

QuickThoughts: https://colab.research.google.com/drive/1AGDwetT6_-CJMSFWv8AUa4ocgRT4JmCl?usp=sharing

MeanSum: https://colab.research.google.com/drive/1AGDwetT6_-CJMSFWv8AUa4ocgRT4JmCl?usp=sharing

In addition, some of the Colabs draw from a GitHub repository where we have made modifications to the code base in some cases (e.g. Python 2 \rightarrow Python 3):

<https://github.com/aaronzr/CS230>

²We had never used containers before, and it took a few days to learn Docker.

8 Contributions

I (Aaron) was originally part of a two-person project group with Ivan Villa-Renteria (ivillar@stanford.edu), who contributed significantly to the proposal and milestone report for this project. Ivan focused more on research and writing, while I took care of implementation on the milestone. He discovered the Skip-Thought and MeanSum papers and had the idea to apply them to our summarization task; meanwhile, I focused on running the algorithms on our own dataset starting from code provided on GitHub.

After the milestone was submitted our team was dissolved since Ivan needed to take an incomplete in CS230. **Thus, all work since the milestone has been a solo effort.** Without Ivan's contributions early on, however, I would have struggled to find an entry point within the vast literature on text summarization, so I thank him for his ideas and partnership.

I would like to especially thank Jo Chuang for his advice, his patience, and his adaptability in a less-than-ideal situation. The help he has given since the end of the academic quarter goes beyond his normal duties as a TA, but it was crucial to the completion of this project and is greatly appreciated.

Finally, I would like to gratefully acknowledge the help of Amelie Byun and Shahab Mousavi in working out a fair resolution to a difficult situation.

References

- [1] Health Resources, Services Administration/National Center for Health Workforce Analysis; Substance Abuse, ' Mental Health Services Administration/Office of Policy, Planning, and Innovation. National projections of supply and demand for behavioral health practitioners: 2013–2025. *HRSA*, 2015.
- [2] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial. *JMIR Ment Health*, 4(2):e19, Jun 2017.
- [3] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth*, 6(11):e12106, Nov 2018.
- [4] Abigail McNathy, Caryn Bello, Debbie J Ellis, Tamara Feldman, Jeffery Binder, Jon Carlson, Katherine Helm, Reid Wilson, Scott Browning, Wes Crenshaw, and et al. Counseling and psychotherapy transcripts: Volume ii. Oct 2015.
- [5] CounselChat. Mental health answers from counselors.
- [6] Nicolas Bertagnolli. Counsel Chat: Bootstrapping High-Quality Therapy Data, October 2020.
- [7] Luís Gonçalves. Automatic Text Summarization with Machine Learning — An overview, November 2020.
- [8] Dr Michael J. Garbade. A Quick Introduction to Text Summarization in Machine Learning, September 2018.
- [9] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. *arXiv:1310.4546 [cs, stat]*, October 2013. arXiv: 1310.4546.
- [10] Ryan Kiros, Yukun Zhu, Russ R Salakhutdinov, Richard Zemel, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Skip-thought vectors. In *Advances in neural information processing systems*, pages 3294–3302, 2015.
- [11] Kushal Chauhan. geniusai-research/email-summarization, October 2020. original-date: 2018-08-01T15:02:10Z.
- [12] Lajanugen Logeswaran and Honglak Lee. An efficient framework for learning sentence representations. *arXiv preprint arXiv:1803.02893*, 2018.

- [13] Eric Chu and Peter Liu. Meansum: a neural model for unsupervised multi-document abstractive summarization. In *International Conference on Machine Learning*, pages 1223–1232, 2019.
- [14] William W. Cohen. Enron Email Dataset.