TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 7921 - 7928

A Deep Learning Approach to overcome the issues in Abstractive Summarization

[1]Dr. Madhavi devaraj PhD,[2]Joel C. De Goma[3]Patricia Chua, [4]LadyLykaDomagsang, [5]Lenard Cledera

[1]Distinguished Professor, [2]Professor [3,4,5]Under Graduate Students

Mapua University, Manila, Philippines

[1]mdevaraj@mapua.edu.ph,madhavidevaraj@gmail.com, [2]jecdegoma@mapua.edu.ph[3]patgeslanichua@gmail.com, [4]ladyjoelyka@gmail.com, , [5]lenardcledera@gmail.com

Article Info
Volume 83
Page Number:7921 - 7928
Publication Issue:
May-June 2020

Article History
Article Received: *19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *18 May 2020*

***Abstract*:**
Generating abstractive summaries has always been difficult since the new words in the summary may disregard the meaning of the text. Abstractive summarization generally faces two major issues: Lack of facts in the summary and repeated words in the summary. This study focused on the challenges of generating novel/nearly accurate summaries. Webis-TLDR-17 corpus is used for this study. The corpus comprises of unstructured social media posts from the social media site, Reddit (2006-2016). In this study, three different models were trained: the pointer generator network (with and without coverage) and the Seq2Seq model as baseline for comparison of the generated summaries. ROGUE Evaluation method was used to calculate the quality of the generated summary. Final generated summaries show the outperformance of the proposed models.

IndexTerms—Attention;Webis-TLDR-17;

1. INTRODUCTION

The Internet opened a lot of possibilities for all the human beings connected to it. Within a few clicks away, you can access a lot of information online and interact with other people from different parts of the world. Discourse regarding a certain topic is easier because of online forums, social media, and electronic news articles with comments section. Due to the increasing volume of data online, one might get discouraged reading through thousands to millions of websites in search of salient information. That's where text summarization comes in.

Text summarization is the process of creating a short, accurate, and fluent summary that contains salient details from a longer text document or multiple documents. The importance of text summarization is to condense long texts to save time and effort in reading them. There are two types of text summarization namely extractive summarization and abstractive summarization. Extractive summarization involves ranking and selecting the most important sentences from the document and creating a summary using the exact original sentences. On the other hand, abstractive summarization involves understanding the document (semantics, words, and sentences) and creating a summary from the linguistics learned from it by generating sentences. It also creates summaries that are closer to what human generated summaries look like because it aims to paraphrase.

This study focused on the challenge of generating novel summaries of unstructured social media posts. Most existing models focus on summarizing structured datasets (Chopra et al. [1], Nallapati et al. [2], Rush et al. [3]) but with the rise of social media, massive amount of unstructured information became available for public as data. A recent study explored the summarization of unstructured text and created a new corpus named Webis-TLDR-17 which contains Reddit[1] posts (Syed, S. [4]). The same study also stated that there are two major problems that are frequently observed in abstractive summarization: 1) inaccurate factual details and 2) repeating words in the summary and suggested to use Pointer-Generator Network to solve these problems (See et al. [5]). In this study, three

[1] http://www.reddit.com

May – June 2020
ISSN: 0193-4120 Page No. 7921 - 7928

different models were trained: the pointer generator network (with and without coverage) and the Seq2Seq model as baseline for comparison of the generated summaries.

This study will take on the challenge of generating novel summaries of unstructured social media posts by using Pointer-Generator Network with Coverage by See et al. [5].

2. LITERATURE REVIEW

Natural language processing (NLP) systems take strings of words (sentences) as their input and produce structured representations capturing the meaning of those strings as their output. Recently, the nature of natural language processing has focused on producing systems which works with human like abilities. There are two problems that are often encountered in this area which are the ambiguity and the complexity of semantic information. NLP has been used to build systems like: spelling and grammar checking, information retrieval, document clustering, information extraction, summarization, text segmentation, machine translation, dialogue systems, etc.

Summarization is simply defined as producing a shorter piece of text (or speech) that captures the essential information in the original [6]. It helps maintain text data by following a set of rules and regulations for efficient usage of text data. Text summarization can be classified on two ways: *abstractive summarization* and *extractive summarization.*

Extractive Summarization techniques generate summaries by selecting a subset of sentences from the original document[7]. It focuses on the important parts of the document based on statistical and linguistic features such as cure words, sentence length, sentence location, term frequency, etc.[8]. Abstractive Summarization methods are of two types: (a) structured-based approach; and (b) semantic-based approach. Structure-based approach encodes most important information from the documents through psychological feature schemas like templates, extraction rules, and various other structures like tree, ontology, and lead and body phrase. Semantic-based approach on text summarization focuses on creating a linguistics illustration of a document or multiple documents which is then fed into a natural language generation (NLG) system.

Deep learning is a branch of machine learning and also a form of a neural network in which many layers of these functions are often chained together. While other machine learning makes predictions based on past observations, deep learning approaches work by learning to not only predict but also to correctly represent the data. Deep learning approaches work by feeding the data into a network that produces successive transformations of the input data until a final transformation predicts the output[9]. Neural networks are a set of algorithms that tries to mimic how the human brain works, which help us group unlabeled data and classify these data if the neural network is trained with a labeled dataset. The major advantage of neural networks is that you can simply "show" it the correct output with the given input [10].

Recurrent neural network (RNN) is a deep neural network that is adapted to sequence data, which makes it an expressive model capable of learning vector-to-vector mappings.

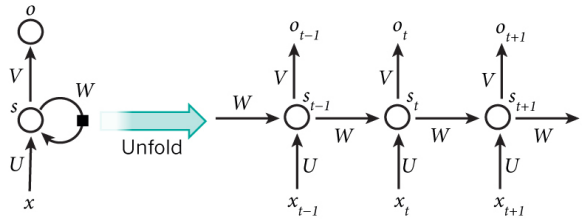

Figure 1. A Recurrent Neural Network Diagram

An example of RNN is an encoder-decoder network, a sequence-to-sequence model. The encoder takes an input and decodes it to a vector. The last hidden state of the encoder gives the encoded vector. The decoder takes the encoded vector and the previous states as inputs and gives the output. To improve this model, *attention mechanism* was developed. Attention is simply a vector proposed a solution to the limitation of the encoder-decoder model encoding the input sequence to one fixed length vector from which to decode each output time step. It is used to relate each word in the output summary to specific words in the input document[11].

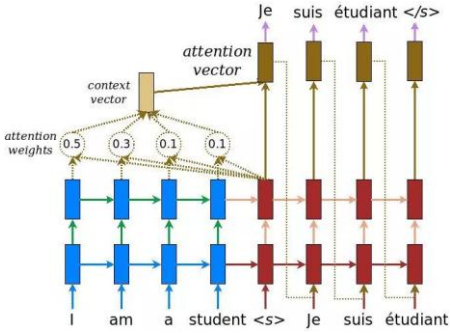

Figure 2. An RNN that uses attention mechanism

Attention mechanism plugs a context vector into the gap between encoder and decoder. The context vector takes all cells' outputs as input to compute the probability distribution of source language words for each single word the decoder wants to generate. With attention, it is possible for the decoder to capture somewhat global information rather than solely to infer based on one hidden state.

Word embeddings are a type of word representation that allows words with similar meaning to have a similar representation [x]. They are a class of techniques where

May – June 2020
ISSN: 0193-4120 Page No. 7921 - 7928

individual words are represented as real-valued vectors in a predefined vector space. Each word is mapped to one vector and the vector values are learned in a way that resembles a neural network, and hence the technique is often lumped into the field of deep learning.

Gradient descent is one of the most popular algorithms to perform optimization and by far the most common way to optimize neural network [x]. Adagrad [x] is a gradient-based optimization algorithm that adapts the learning rate to the parameters, performing smaller updates (i.e. low learning rates) for parameters associated with frequently occurring features, and larger updates (i.e. high learning rates) for parameters associated with infrequent features. For this reason, it is well suited for dealing with sparse data.

Nallapati, R., et. al. (2016) [2]proposed a model for abstractive text summarization using Attentional Encoder-Decoder Recurrent Neural Networks. The study applied the off-the-shelf attentional encoder-decoder RNN to summarization and showed that it already outperforms state-of-the-art systems. They proposed several novel models, including an encoder-decoder model, each addressing a specific weakness in the baseline. The attentional encoder-decoder yielded very promising results. Each of their proposed novel models addresses a specific problem in abstractive summarization, yielding further improvement in performance.

Syed, S. (2017) [4]created a novel, usable dataset from the domain of social media called the Webis-TLDR-17 corpus. It is derived from a large set of Reddit posts spanning ten years. They mentioned that one of the key challenges for neural networks in dealing with language is understanding unstructured/informal text. Properly written, domain-specific texts like news articles are usually not enough to extensively explore the capabilities of an automatic summarization system. Syed argues that a dataset from Reddit, a social media platform, where users communicate informally can greatly help the research community. Syed cleaned and extracted a dataset specifically suited for abstractive summarization. Reddit is a social media platform wherein users can submit content and concerned users can comment on said content. Submissions are usually large texts discussing in depth about a topic, while the comments posted by readers are usually terser. Nevertheless, both submissions and comments can contain a summary of the content, written after the abbreviation tl;dr (too long, didn't read). The tl;dr text can be regarded as the summary of the whole post and use this pair of content-summary as an input to train a machine learning algorithm.

Syed suggested See, A., et. al. (2017)'s proposed summarization model [5]called the Pointer-Generator network. This approach deals with two key problems on abstractive text summarization: (1) failure to reproduce factual details; and (2) word repetition. First, the researchers used a hybrid pointer-generator network that can copy words from the source text via pointing, while retaining the ability to produce novel words through the generator. Second, they used coverage to keep track of what was summarized which served as a solution of one of the text summarization problems on repetition.

ROUGE or Recall-Oriented Understudy for Gisting Evaluation is a set of metrics that was introduced to give a score based on the similarity in the sequences of words between a human-written model summary and the machine summary. It helps automatically evaluate machine-generated summary. It includes five measures like ROUGE-N, ROUGE-L, ROUGE-W, ROUGE-S and ROUGE-SU [x]. Although for abstractive text summarization, Syed suggested that ROUGE metric is not a suitable measurement for accuracy. Their study much preferred human evaluation like the Document Understanding Conference (DUC) Linguistic Quality Questions [x]. All linguistic quality questions require a certain readability property to be assessed on a five-point scale from "1" to "5", where "5" indicates that the summary is good with the respect to the quality under question, "1" indicates that the summary is bad with respect to the quality stated in the question, and "2" to "4" show the gradation in between. The questions are: grammaticality, non-redundancy, referential clarity, focus, and, structure and coherence.

3. RESEARCH METHODOLOGY

3.1 Data Gathering

The researchers used the Webis-TLDR-17 corpus from[4]. The data from this corpus comes from the social media site, Reddit (2006-2016). It contains 3,848,330 posts with an average length of 270 words for content, and 28 words for the summary. To improve the dataset, the researchers set the following standards:

- the content should at least contain 100 words; and
- the reference summary (tldr's) should at least contain 10 words.

After checking the lengths of the texts, the dataset contained a total of 345, 840 articles. These were split into a 90-5-5 ratio for the training (311, 256 articles), validation (17, 292 articles), and evaluation (17, 292 articles) sets.

3.2 Data Processing

The data processing phase began by extracting the "content" and "summary" field from each text file from the corpus. Next, sentences were separated by line using python functions and then the data was tokenized using the Stanford Tokenizer[12]. The data was converted into lowercase and is made into a single string enclosed with <s> and </s> tags. Data is read from the saved files and

written to serialized binary files: train.bin, test.bin, and val.bin. The data was chunked into manageable parts of 1000 examples per chunk.

3.3 Model Training & Validation

The researchers trained three different models: the pointer generator network (with coverage) model, the pointer generator network (without coverage) model, and Nallapati's model(for comparison). In order to replicate See's method, the researchers ran a concurrent validation script along with the training script. Training for each model took approximately 3 and a half days. The researchers used Crestle.ai to host the study's files, and for the training and the validation phase on a high-performance GPU NVIDIA Tesla P4 GPU. Crestle.ai is a platform for quick AI deployment. It is a convenient tool for testing neural networks that bills at $2.40/hour.

3.4 Model Testing

Testing phase began with decoding the test dataset using the models. The decoding python script used beam search of size 4, maximum decoding tokens of 100, and minimum decoding tokens of 10, to produce the summaries and placed them in separate folders: decoded and reference.

After generating the decoded summaries, testing the model required the study to use the ROUGE tool. The researchers used pyrouge a python wrapper for the ROUGE summarization evaluation package. Scripts using pyrouge were created and ran in the terminal to be able to generate the average ROUGE scores of all the articles in the test data set. The precision, F1, and recall scores of the ROUGE-1, ROUGE-2, and ROUGEL metrics were computed.

The ROUGE scores alone are not enough to test the model, so the researchers sought out an expert to answer a survey with 5 text files, wherein each text/post had 4 corresponding summaries from: Nallapati's model, pointer-generator model, pointer-generator with coverage model, and the actual reference summary. The summaries were manually evaluated using the DUC Linguistic Quality Questions[13]. This metric included five readability properties: Grammaticality, Non-redundancy, Referential clarity, Focus, and Structure and Coherence. Each property were assessed on a five-point scale from "1" to "5", with "5" being the highest.

3.5 Prototype Building

The prototype is built on an Angular 6 platform with Flask as the micro framework. The researchers first built three API's that calls processing data, decoding text, and rouge evaluation separately. The web application also uses the same packages mentioned above, Stanford Tokenizer and Pyrouge, for processing the data and evaluating the generated summary respectively. For the front-end, the prototype used Angular Material for its components. The prototype is designed to have the ability to run on the three different models that the researchers trained, as well as evaluating the generated summaries through ROUGE evaluation.

4. RESULTS AND DISCUSSION

For reference, this and the succeeding chapters will use the keywords: PGen, PGenCov, and Seq2Seq to refer to the pointer generator network model, pointer generator network (with coverage) model, and the sequence-to-sequence attentional model respectively.

4.1 "Good" Summaries vs. "Bad" Summaries

The proponents of this study classified the results (summaries) as "good" and "bad" based on the information it retained and its faithfulness to the source material. Any summary that did not have any information from the source was considered a "bad" summary and any summary that holds relevance to the source was considered a "good" summary. This is not a rating of the summaries, this is merely a term used for the presentation of the results of the model.

4.2 Dataset

After data processing, the dataset was left with posts from categories such as dating, life, advice, gaming, and various tutorials. These posts commonly asks for advice and/or assistance. The model fixated on this format which is why when decoding the test data set, articles which contained questions with "how…?" produced "good" summaries.

4.3 ROUGE Evaluation Results

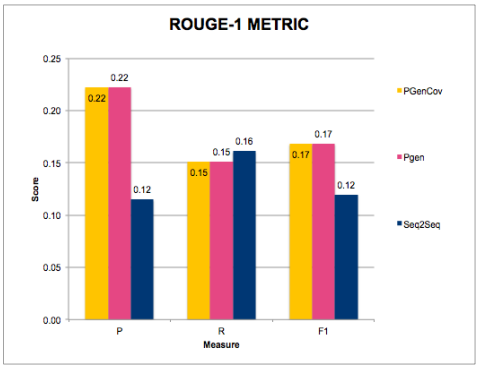

Figure 3. ROUGE-1 metric for PGenCov, PGen, Seq2Seq models. PGenCov and PGen both have metrics of 0.22, 0.15, and 0.17 for precision, recall, and F1 measures respectively. Seq2Seq has 0.12, 0.16, and 0.12.

May – June 2020
ISSN: 0193-4120 Page No. 7921 - 7928

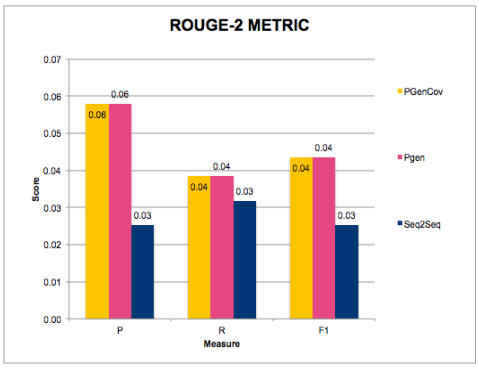

Figure 4. ROUGE-2 metric for PGenCov, PGen, Seq2Seq models. PGenCov and PGen both have metrics of 0.06, 0.04, and 0.04 for precision, recall, and F1 measures respectively. Seq2Seq has 0.03 in all measures.

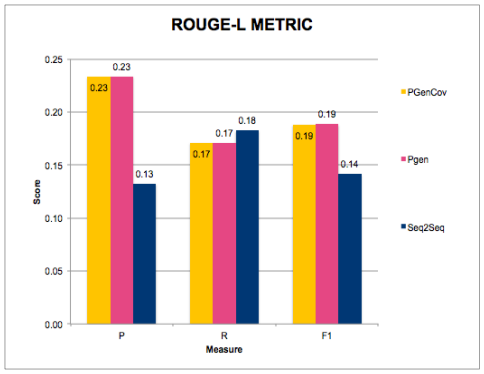

Figure 5. ROUGE-L metric for PGenCov, PGen, Seq2Seq models. PGenCov and PGen both have metrics of 0.23, 0.17, and 0.19 for precision, recall, and F1 measures respectively. Seq2Seq has 0.13, 0.18, and 0.14

Figures 1-3 show that although all of the models scored poorly, both the PGen and PGenCov still outdid the Seq2Seq model. A problem that arose during the experiment was working on data that the researchers had to repeatedly clean. As a result, the amount of data kept decreasing, and it may have been a possible cause of the poor summaries.

4.4 Survey Evaluation Results

A survey was answered by a professional freelance literary agent wherein they gave ratings on the result of the three models' and the reference summary. These ratings were based from the DUC linguistic quality questions: grammaticality, non-redundancy, referential clarity, focus, and structure and coherence. Figure 4 displays the average rating of the model-generated and reference summaries.

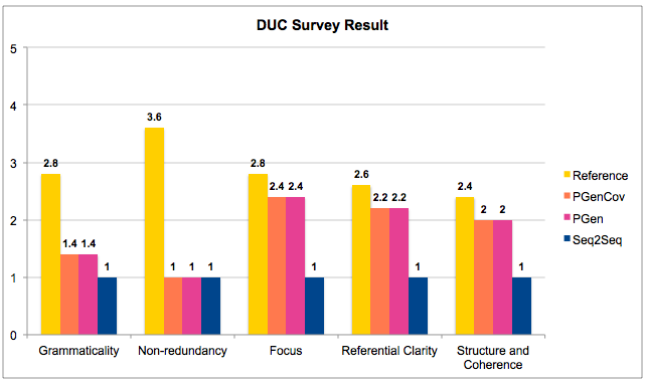

Figure 6. The average rating of the reference summary and the three system-generated summaries. The ratings are based on the DUC Linguistic Quality Questions: Grammaticality, Non-redundancy, Focus, Referential Clarity, and Structure and Coherence.

This bar chart shows that the average rating for the reference (human-generated) summary is close with the ratings for the PGen/PGenCov models. As shown in this chart, 56 the average rating for the human-generated summaries are the highest, followed by PGen/PGenCov models and the one with the lowest score is the Seq2Seq model. The survey shows, same with the ROUGE evaluation, that despite the low ratings, the PGen/PGenCov (proposed model) scored higher than the Seq2Seq model (baseline model) and it is much closer (in ratings) with the reference summaries.

Based on the results, the PGen and PGenCov models both outdid the Seq2Seq model and both have similar ratings with the reference summary. All three models have the lowest rating on redundancy which shows that the coverage was not able to solve the issue on repeating words. Focus and referential clarity have similar ratings between the reference and the PGenCov/PGen summaries, which says that the proposed model was able to preserve factual and relevant information from the post.

Below is a sample survey question with comments from the expert. The text inside the parentheses refer to the annotations done.

Text Post: Hi there, I'm currently making my own LED screens for my bands live production and i'm having a bit of trouble sourcing the right parts. I need to find a suitable plug and socket for the loom of cables going between the drivers (LED controller) and the screens themselves. There is 20 6 metre long 18AWG wires in each loom for each screen. I need to be able to plug the cables in and out of the screens very quickly (during setup and setdown). This is definitely the closest thing I've found but I'm worried they may be a bit bulky for my screens which are very narrow and long. [URL] Does anyone know if there is a smaller, maybe cheaper version of these kinds of plugs? Any help is appreciated.

May – June 2020
ISSN: 0193-4120 Page No. 7921 - 7928

reference summary: need to find the right crimp plug and socket for cables i'm making for an led screen project [examples in post] *("There was no specific subject or doer of the actions. However, this is still found as relevant to the main topic of the post. This could have been written as, "What are the essential parts/equipment to be used in making LED screens for band production?"")*

PGenCov&PGen summary: i need to find a suitable plug for my bands, but i don't know what to do. i don't know what to do. *("This may be relevant to the post, but the repetition of words was still noticeable and unnecessary.")*

Seq2Seq summary: i [21m] have a girl [21f] of a year, and i don't know if i don't know if i don't know if i don't know if i don't know if i don't know if i don't know what to do. *("This could have just been eliminated as the idea was not even related to the initially posted text.")*

Table 1. Sample DUC rating evaluation for Grammaticality, Non-redundancy, and Focus

Summary	Grammaticality	Non-redundancy	Focus
Reference	2	3	2
PGenCov	1	1	2
PGen	1	1	2
Seq2Seq	1	1	1

Table 2. Sample DUC evaluation rating for Referential Clarity and Structure and Coherence

Summary	Referential Clarity	Structure and Coherence
Reference	2	2
PGenCov	2	2
PGen	2	2
Seq2Seq	1	1

4.5 Issues PGenCov encountered

4.5.1 Word Embeddings

Due to the various topics of the dataset, the "closeness" of words and the limited vocabulary affected the outcome. Word embeddings were created from the dataset itself instead of using a pre-trained one which in result lead to the model using a bounded set of words.

4.5.2 Pointer-Generator

Factual information wasn't preserved as expected since spelling errors greatly affected how the words related to each other. Some reference summaries were also unreliable, and some articles were written in other languages aside from English, which made the relation of words difficult for the model.

4.5.3 Coverage

Errors tend to occur when the attention is more scattered, indicating that perhaps the network is unsure todo. In some cases, the model could not produce a relevant summary, which is why it generated the words with the highest attention probability, which for this study, was the phrase: "i don't know".

4.6 Generated Summaries

The following are sample generated "bad" and "good" summaries from the three models: PGen, PGenCov, and Seq2Seq.

4.6.1 Good Summaries

4.6.1.1 PGen

text post: It's been so long since I've had an in depth conversation about anything that I'm starting to forget what it felt like. I remember college classes debating topics and what not, but when it comes to day-to-day conversations it feels like everyone is literally only making small talk with me, even my few friends and my family, like we've already talked about everything and there's not much new anymore. When i try to make new friends, it's always superfluous chit-chat for a little while that quickly drops off into nothing. So my question is two fold - one, how can i make small talk more interesting? and two, how do you turn small talk into real conversation?

reference summary: i'm kind of starting to feel like i don't know how to hold a conversation anymore.

generated summary: how do i make small talk to dayto-day talk to my family? i don't know what to do with my family, but i don't know what to do. i don't know what to do.

4.6.1.2 PGenCov

text post: i know this is a controversial opinion, but i'm not a big fan of the games dialogue and humor. however the addicting gameplay and art style is more than enough for me to label BL2 as my favorite game to date. i recently bought it on steam, just cause i wanted to replay it with Physx, and have a friend who owns a copy of the game too. he's never beaten a single playthrough of the game because the jokes are so cringy, and although i agree with his opinion on that i personally think it's a good enough game to look past that. how can i convince him to do the same?

reference summary: how can i convince a friend to look past the dialogue and humor in this game?

generated summary: how can i convince my favorite friend? i'm not sure how to do with my favorite game, and i don't know what to do.

4.6.1.3 Seq2Seq

Because the model performed so poorly, the researchers were not able to find a "good" summary" that was generated.

May – June 2020
ISSN: 0193-4120 Page No. 7921 - 7928

4.6.2 Bad Summaries

4.6.2.1 PGen

text post: hey everyone, i managed to get stuck on a planet and i dont know what to do anymore. so i found another ship and tried to repair that. then my game crashed and when i logged back in my old ship was gone. no i only have a broken ship on a planet without the stuff to fix it. there is no zinc anywhere. no yellow flowers. i also font have any hear with zinc in it salvage. i also cant find a trading post, im wandering around since two hours. any ideas? sadly i dont have a save game or anything before this. if only i could start with just the starting engine. i can see the space port from whemim at…

reference summary: sos, stuck on planet.. stuck on planet. no zinc. tried everything help. sos. edit: made it, thanks [found a trade post after ~ three hours]

generated summary: i don't know what to do. i don't know what to do. i don't know what to do. i don't know what to do.

4.6.2.2 PGenCov

text post: so starting this year, we are getting a mechanical engineering minor, hopefully in the next year or so it'll be a full major. so i took a few cad classes in high school too, and i loved them. for the most part of what i've found, most cad classes transfer to be different "design/3d-modelling" in the computer engineering, which i have no idea how much is even related to things similar to cad.as for freshman dorms, president's park is a big ring of housing, it's nice and traditional, but the plumbing and most commodities are out dated. the commons was renovated a year or two ago, and it's brand new, had its air conditioning on before the president's park, but they won't be having trilets this year from what i've heard.

reference summary: CAD will most likely fall under computer engineering, mechanical engineering is a minor that will hopefully become a major soon, live in the commons if comfort is a big issue, and gaming at mason is prety easy to fit into.

generated summary: i'm not sure if i don't know what to do. i don't know what to do. i don't know what to do.

4.6.2.3 Seq2Seq

text post: so about 6 months back, a friend of mine sent me an email explaining some things he was unhappy about and telling me that he wasn't cutting me out of his life, but he needed some space to calm down and think about things. i waited a couple of weeks and then emailed him back, apologizing for hurting his feelings and saying that i wish he had brought up these issues earlier. months later i realize he has completely blocked me on facebook (which he made no mention of). not only this, but he also blocked my SO. his SO and i have been hanging out and talking perfectly noremally in the meantime (albeit a bit less often than usual). we're currently trying to have a conversation and hash everything out and i have a problem. i feel like him blocking me was an escalation and a sign that he did not want me in his life anymore. i feel very strongly that him blocking my SO counts as involving him in our problems and completely inappropriate. my friend thinks that blocking my SO without a word was the opposite of involvement and that if i'm upset "that's really a [me] problem". am i overreacting? is he being dismissive of my feelings?

reference summary: me [26f] with my friend [26m] of 8 years. i feel like he escalated an argument. he feels like i'm overreacting. my friend of 8 years told me that he needs some space and then blocked my SO and i on facebook. i feel like this is an unwarranted for escalation and he thinks this is a "me" problem. am i overreacting or is he being dismissive of my feelings?

generated summary: my [21m] girlfriend [21f] of 2 months, she isn't a relationship. my boyfriend isn't a lot of my friend, and i don't know what to do.

As seen in the above examples, the summaries generated by the PGenCov is fixated on some few phrases, one of which is " I don't know what to do". The reason why this happens is, "I don't know what to do" phrase has the highest probability of generation. They are the most frequent words used together in the corpus. Since reddit is a website where people generally ask for advice, this is understandable to happen.

Given that the dataset was reduced in size due to the requirements of this study (minimum number of words, minimum number of words in the title/summary, etc) the model was focused on the words used on the remaining dataset.

While coverage was supposed to remove redundant phrases, it only covered bigrams, meaning that it only checks words by pairs, which is why redundancy commonly happens after the 3rd word. The researchers proposed that coverage includes trigrams in its computation for the coverage vector. In the example of good summaries, it shows that the pointer-generated summary can correctly identify the general idea of the article. It does not simply do an extraction of words from the source article, it was able to generate its own words even though it was only a single sentence.

Due to having a small(er) dataset, the vocabulary of this corpus was limited to the topics under it. For example, photography and music falls under the same category of media. What happens is that, words generated for an article about photography may include words related to music, because according to the word embeddings they have the closest relationship to each other compared to the other words.

The researchers recommend using a dataset with one single topic or related topics so that the generated words of this model will match the subject of the article and will not have issues with regards to generating unrelated words. A large dataset for unstructured data has yet to be gathered that can fit in the requirements for testing a summarization model.

5. CONCLUSION

This paper presents an experiment on applying an unstructured dataset to a text summarization model that gave promising results with news articles as inputs. Though the PGen/PGenCov models produced great results with

May – June 2020
ISSN: 0193-4120 Page No. 7921 - 7928

news articles as input, it didn't generate the expected results with the social media posts. Also for this study, redundant phrases mostly occurred after 3 words which usually happened when the model ran out of words to generate. The coverage mechanism couldn't help the model produce a relevant summary, and repeatedly produced the phrase: "i don't know", because the attention was too scattered. In terms of factual information, some of it were preserved, but a lot of factors greatly affected the preservation like spelling errors and the unreliability of the reference summaries.

REFERENCES

1. S. Chopra, M. Auli and A. M. Rush, "Abstractive Sentence Summarization with Attentive Recurrent Neural Networks," in NAACL-HLT 2016, San Diego, California, 2016.
2. R. Nallapati, B. Zhou, C. dos Santos, C. Gulcehre and B. Xiang, "Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond," in Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL), Berlin, Germany, 2016.
3. A. M. Rush, S. Chopra and J. Weston, "A Neural Attention Model for Sentence Summarization," in Proceedings of the 2015 Conference of Empirical Methods in Natural Language Processing, Lisbon, Portugal, 2015.
4. S. Syed, Abstractive Summarization of Social Media Posts : A Case Study Using Deep Learning, 2017.
5. A. See, P. J. Liu and C. D. Manning, "Get To The Point: Summarization with Pointer-Generator Networks," 2017.
6. A. Copestake, "Natural Language Processing," 2004. [Online]. Available: https://www.cl.cam.ac.uk/teaching/2002/NatLangProc/revised.pdf. [Accessed 5 January 2018].
7. M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. Trippe, J. B. Gutierrez and K. Kochut, "Text Summarization Techniques: A Brief Survey," Georgia, 2017.
8. N. Bhatia and A. Jaiswal, "Trends in Extractive and Abstractive Techniques in Text Summarization," International Journal of Computer Applications, vol. CXVII, no. 6, pp. 21-24, 2015.
9. Y. Goldberg, "Neural Network Methods for Natural Language Processing," in Synthesis Lectures on Human Language Technologies, Toronto, Morgan & Claypoo, 2017, pp. 1-12.
10. P. McCollum, "An Introduction to Back-Propagation Neural Networks," Encoder, 2009. [Online]. Available: http://www.seattlerobotics.org/encoder/nov98/neural.html. [Accessed 6 January 2018].
11. J. Brownlee, "Attention in Long Short-Term Memory Recurrent Neural Networks," 30 June 2017. [Online]. Available: https://machinelearningmastery.com/attention-long-short-term-memory-recurrent-neural-networks/. [Accessed 20 January 2019].
12. C. Manning, T. Grow, T. Grenager, J. Finkel and J. Bauer, "PTBTokenizer," [Online]. Available: https://nlp.stanford.edu/software/tokenizer.html. [Accessed 20 March 2019].
13. L. Buckland, "D U C 2 0 0 7: Task, Documents, and Measures," 24 March 2011. [Online]. Available: https://duc.nist.gov/duc2007/tasks.html. [Accessed 7 July 2018].