Adaptation of text to publication type

Luka Žontar
University of Ljubljana,
Faculty of Computer and Information Science
Ljubljana, Slovenia
zontarluka98@gmail.com

Zoran Bosnić
University of Ljubljana,
Faculty of Computer and Information Science
Ljubljana, Slovenia
zoran.bosnic@fri.uni-lj.si

ABSTRACT

In this paper, we propose a methodology that can adapt texts to target publication types using summarization, natural language generation and paraphrasing. The solution is based on key text evaluation characteristics that describe different publication types. To examine types, such as social media posts, newspaper articles, research articles and official statements, we use three distinct text evaluation metrics: length, text polarity and readability. Our methodology iteratively adapts each of the text evaluation metrics. To alter length, we focus on abstractive summarization using text-to-text transformers and distinct natural language generation models that are fine-tuned for each target publication type. Next, we adapt polarity and readability using synonym replacement and additionally, manipulate the latter by replacing sentences with paraphrases, which are automatically generated using a fine-tuned text-to-text transformer. The results show that the proposed methodology successfully adapts text evaluation metrics to target publication types. We find that in some cases adapting the chosen text evaluation metrics is not enough and we can corrupt the content using our methodology. However, generally, our methodology generates suitable texts that we could present to a target audience.

KEYWORDS

text adaptation, context-aware, artificial intelligence, text summarization, natural language processing

1 INTRODUCTION

With more and more internet usage, the textual data on the internet is highly increasing. However, different media target different audiences and thus an arbitrary article may not be appropriate for everyone. Consequently, already published content is being rewritten and adapted for other target audiences.

Why is targeting audiences so important? When speaking with someone in person, we adjust body language, tone and the words we use, so that the audience understands the message we are trying to send. In a similar manner, we also have to be aware of the target audience when writing. Even though the task of adapting texts to different audiences may look easy to experienced writers, rookies and amateurs may struggle in selecting the information that might be relevant to a particular target audience. Nevertheless, a way to deal with words and some common sense should be enough to complete the task, but due to the latter requirement automating this task becomes a much harder problem.

Information Society 2020, 5–9 October, 2020, Ljubljana, Slovenia

In this paper, we adapt texts to context by manipulating three text evaluation metrics: length, polarity and readability. Our method will be able to transition between social media posts, research articles, newspaper articles and official statements, where each publication type targets a different audience. While governmental institutions and academics both publish neutrally-oriented texts, research articles tend to be much longer than official statements. Social media and news usually target wider audiences, which is why texts should be more readable. However, the two can be separated by the amount of opinion we can include. Newspaper articles should be less biased and thus include less positively or negatively-oriented words.

Our methodology iteratively adapts key text evaluation metrics towards the mean values of the target publication type that will be calculated from a sample set of articles. In each iteration our method first manipulates length using abstractive summarization techniques and natural language generation models. Next, it replaces words with more appropriate synonyms and adjusts polarity and readability scores. Finally, it uses a fine-tuned text-to-text transformer to generate more appropriate paraphrases that replace whole sentences in our text and alter readability.

2 RELATED WORK

A lot of research has already been done on how to evaluate and alter text and we will use many existing methods to help us develop our methodology. Kiefer [6] in her article describes many text evaluation metrics such as the percentage of abbreviations and lexical diversity in text. However, how can one alter the percentage of the percentage of abbreviations meaningfully? We picked three text evaluation metrics that can be reasonably altered using existing methods. Flesch [4] developed an equation that determines the readability of the text using the number of words per sentence and the number of syllables per word ratios. Even though structure-based metrics are important, we also have to consider the message of the text. Using sentiment analysis, we can determine whether the writer has positive or negative affections towards the topic of the text. Feldman [3] in his article discusses several approaches of sentiment analysis based on the unit that we will be classifying (i.e. documents, sentences, aspects).

As length is one of the chosen text evaluation metrics that we wish to adapt, we have to be able to both summarize and extend the text. According to Allahyari et al. [1], we differ extractive and abstractive summarization approaches. Extractive approaches shorten the original text by excluding less relevant sentences. Significancy of the sentence can be evaluated by determining whether the sentence is related with the main topic or whether its content is distinctive in comparison to other sentences. On the other hand, abstractive approaches tend to summarize texts in a new (more human-like) manner by structuring the text into some logical form such as graphs, trees and ontologies [5] or by consider text sentiment and constructing Semantic graphs [7].

When adapting shorter texts to longer, natural language generation has proven to be a very strong tool. Radford et al. [8] developed a natural language generation technique to generate additional text and produced state of the art results using unsupervised multitask learners for model learning. Their model was trained to predict the next word in text based on 40GB of Internet content. They concluded that large training datasets and models trained to maximize the likelihood of a sufficiently varied corpus can learn a surprising amount of tasks, while no supervision is needed in training.

Another method that is commonly used when adapting texts to context is paraphrasing, i.e. rewording of something written by changing its structure or replacing the words with their synonyms. Goutham in his article [10] used a pre-trained text-to-text transfer transformer to generate paraphrases of questions. The model was fine-tuned, where the input texts were questions from Quora and the expected output were the questions that were labeled as their duplicates.

In our paper, we plan to exploit the aforementioned abstractive summarization technique to shorten our texts and fine-tuned the pre-trained natural language generation model that Radford et al. [8] developed. Similarly as Goutham [10], we intend to fine-tune a pre-trained text-to-text transformer that would be able to generate paraphrases of a sentence. To calculate readability score of the input text, we plan to use the formula proposed by Flesch [4].

3 ADAPTATION OF TEXT

As mentioned before, the proposed method iteratively manipulates the chosen text evaluation metrics to adapt text to different target audiences. In Figure 1, which gives an overview of the method, we can see that before we start running the process, we calculate the initial values of text evaluation metrics for each publication type as the average values of a set of documents. We used dataset of 150 documents for each publication type, where all the documents hold text that contain COVID-19 related content, with which we minimize the effect of variables that we will not take into account in text adaptation. We also define the number of iterations (in our case: 5) and the acceptable error ϵ (in our case: $\epsilon = 0.1$) that determines whether it is still worth altering a particular text evaluation metric.

In each iteration, relative differences between current and initial values of text evaluation metrics are calculated. If the absolute relative difference to some metric is bigger than ϵ, we try to adjust it to the targeted value. We adjust key text evaluation metrics in the main loop of the process in Figure 1 using the following procedures:

- In case the target length is smaller than its current value, we use a pre-trained **T5 text-to-text transformer [9]** to **summarize** the input text. The model is an encoder-decoder model that uses transfer learning on a model that is firstly pre-trained on a data-rich task and then fine-tuned on a downstream task.
- To generate additional text, if the input text is shorter than the average text of the target publication type, we use **fine-tuned natural language generation models**. A pre-trained GPT2 natural language generation model [8] that is based on the aforementioned unsupervised multitask learners is fine-tuned on documents of each considered publication type. We get four distinct NLG models, where each generates text similar to the ones that it was fine-tuned on. Consequently, we would assume that the generated text needs less further adaptation.
- While adapting length might be the procedure with the most visible results, we also have to adapt the other text evaluation metrics. We develop a **synonym replacement** procedure to adjust polarity and readability scores to the target values. The procedure is executed in iterations and in each iteration we replace the word with the highest sum of absolute relative differences of polarity and readability scores to the initial values of the target publication type with its optimal synonym, i.e. the synonym which causes the sum of absolute relative differences to minimize.
- Finally, we alter readability by **generating paraphrases with a T5 text-to-text transformer [10]** that was fine-tuned to generate paraphrases by learning on Microsoft Research Paraphrase Corpus dataset [2]. We then pick the optimal paraphrase, which minimizes the relative difference to the target readability score.

Replacing sentences with their paraphrases could potentially also alter length and polarity. We test the assumption by generating five paraphrases for each sentence in 100 documents for each considered publication type and find that the relative difference of length and polarity between the initial sentence and its paraphrases is not significant. The obtained mean relative difference of polarity scores in this preliminary analysis was 0.91e−3 and the mean relative difference of lengths was 0.11e−3.

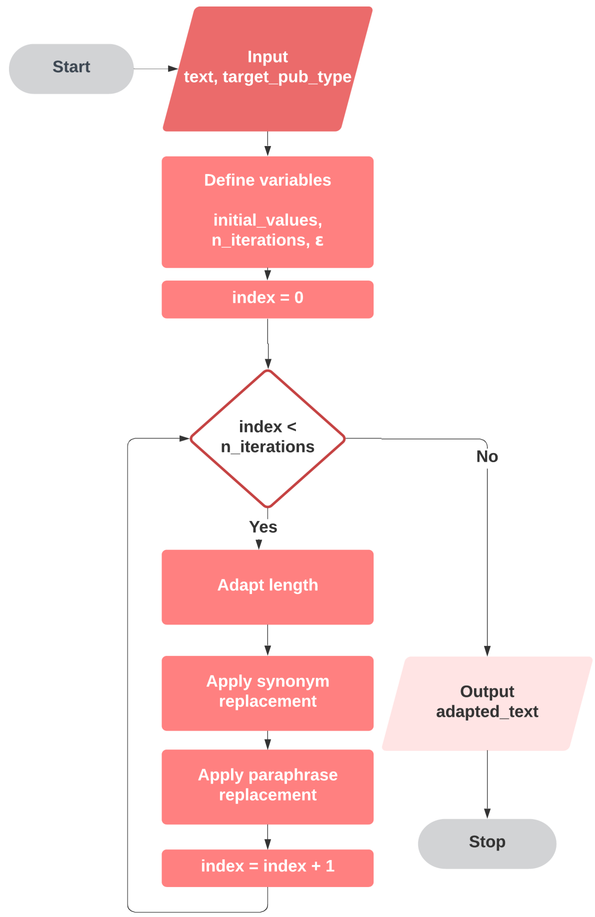

Figure 1: Flowchart of the text adaptation methodology

4 EVALUATION AND RESULTS

In our experiments, we evaluate the quality of text transformation between all possible pairs of four different publication media types: social media, news, research articles and official statements. We tested our methodology by generating adapted texts of a subset of texts used for calculating initial values that consists of 100 documents for each publication type (i.e., 400 altogether) that were randomly chosen from the main dataset. We adapted each document to the other three publication types and thus test all of the 12 possible transitions. We observed how the key text evaluation metrics behaved and whether the generated text was meaningful or not. The results text evaluation metrics before and after adaptation to context are shown in Table 1.

From Table 1 we can observe, that the text evaluation metrics successfully changed in the right direction. In most cases we significantly improved the values of metrics. The length manipulation managed to consistently decrease the relative difference towards targeted length and in many occasions even converge under ϵ value. Polarity and readability scores seem harder to adapt. However, in each case we successfully adapted the sum of relative differences of those metrics, with which we can conclude that synonym replacement method performs suitably, too. Its inefficiency may be caused by the lack of choice in synonym and paraphrase replacement and the limited amount of words and sentences that can be replaced. In the following, we provide two examples of the content that was generated with our methodology.

For example, if we try to adapt social media post to a newspaper article, we can notice in Figure 2 that a very imaginative paraphrase was generated. However, the start of the paraphrase "Under the new law, you no longer have to wear" means exactly the opposite as the input text message suggests. If we exclude that fact, the generated text looks a lot like a newspaper article. Text generated with NLG model is related to the content and well formed. Overall, the introduction that is shown in Figure 2 looks like an exemplary introduction to a newspaper article.

You now have to wear face coverings when out shopping - this includes a visit to your local Community Pharmacy…

Under the new law, you no longer have to wear face coverings when out shopping, especially if you have to take a trip to a local community pharmacy. The Canadian Press (@cp) may 4, 2016 - A new study suggests the number of COVID-19 cases in greater sudbury has dropped by more than half since coronavirus testing began two weeks ago at community pharmacies across canada. But some health professionals worry that too many people are still buying items like masks and gowns without proper gloves or protective gear because they don't feel safe putting them on before heading into stores today.

Figure 2: Example of text adaptation from the social media post to the newspaper article

Switching to another transition to adapt a newspaper article to a social media post, we generate a text that is presented in Figure 3. It talks about the first 1,000 cases in Canada. The sentence "Canada is officially in" indicates that Canada is a part of something, but the word that continues this sentence in the original text is "recession". This significantly changes the meaning the text. However, we could also interpret this sentence by saying that it joined the countries with more that 1,000 cases. Additionally, it is hard to understand the first part of the second sentence, but we do conclude with a very informative sentence about the COVID-19 state in Canada that we could publish on Twitter.

Figure 4 shows an example of transforming a tweet to an official statement. The method generates "Authentics: Politians

Canada is officially in. Their jobs due to more than 1,000 state of mattereous state filled - roughly 6,000 tests and 11 confirmed cases of COVID-19 a report also details an estimated 79.

Figure 3: Example of text adaptation from the newspaper article to the social media post

and bussinesses want a story like this" from "Politicians and bussinesses want your money". While the paraphrase is not the most accurate, it connects well with the rest of the text. Replacing "bad" with "unspeakable" suggests that the original text either has higher readability or polarity score than the initial values of social media posts, because the word is usually used as more negatively-oriented, i.e., it decreases polarity, and it contains more syllables, i.e., it decreases readability. The generated official statement is connected to the input text and is much more appropriate for a governmental institution to publish than the input sequence.

#coronavirus #covid19 deaths continue to rise. It's almost as bad as it ever was. Politicians and businesses want your money.

#coronavirus #covid19 deaths continue to rise. It's almost as unspeakable as it ever was. Authentics: Politicians and businesses want a story like this. Co/2jh6ocm7vj0 — rt (@rt_com) february 2, 2017 Dr. Sanjay Gupta of CNN medical center in New Delhi says the virus has spread through air across Assam: "more than 1,200 peoples have been hospitalized with respiratory light-headednesses or severe skin rashes after being exposed". (source.) His hospital reports that there are about 937 cases confirmed by them so far for every 100500 patients treated at his facility. . . The WHO is currently monitoring two additional countries – nepal(with 1076 radical case count from march 15-17)and bangladesh (1116). We believe these numbers may need further updates when they become available," said Dr. Sushil Kumar Panicker, director general & chief medical officer of world health organization …

Figure 4: Example of text adaptation from the social media post to the official statement

While our method successfully adapts key text evaluation metrics, our results are not perfect when it comes to the content. Our method has its drawbacks such as generating lots of additional content, which often results in an unconnected text. Additionally, synonym replacement and paraphrase generation can incorrectly replace original sentence or word, where the paraphrase or synonym changes the meaning but proves to be effcient when adapting text evaluation metrics, if there exist such synonyms that are more appropriate to use for a particular target audience. Nevertheless, our methodology generated a few sequences that could be published for target audiences without any changes and lots of texts would only require minor corrections.

To conclude this section, we are satisfied with the benchmarking results that our method produced in adapting key text evaluation metrics. The methodology produces some interesting content and can thus be used as a baseline for further text adaptation to target audiences.

5 CONCLUSION

In this article we developed a methodology that adapts texts to context. The methodology focuses on three text evaluation metrics: length, readability and polarity of the text. Our method iteratively adapts text to the calculated initial values based on the targeted publication type by adjusting the key text evaluation metrics. We successfully managed to adjust text evaluation metrics in nearly all transitions.

While we found text evaluation metrics that define different publication types, in some cases adjusting these measures is not

Input publication type		Target publication type: Official statements		Research articles		News		Social media	
		Initial	Adapted	Initial	Adapted	Initial	Adapted	Initial	Adapted
Official statements	Length			0.79	0.04	0.04	0.03	36.39	0.35
	Polarity			2.88	0.15	2.05	0.04	2.78	0.4
	Readability			0.36	0.75	0.23	0.35	0.4	0.24
Research articles	Length	3.06	0.05			2.99	0.04	136.23	0.33
	Polarity	0.81	0.27			0.33	0.07	0.18	0.46
	Readability	0.17	0.08			0.34	0.22	0.45	0.12
News	Length	0.97	0.03	0.99	0.03			63.79	0.4
	Polarity	0.88	0.14	0.43	0.1			0.33	0.37
	Readability	1.21	0.05	1.2	0.84			0.24	0.11
Social media	Length	0.69	0.02	0.64	0.03	0.97	0.04		
	Polarity	0.85	0.28	0.28	0.02	0.55	0.06		
	Readability	0.71	0.27	0.69	0.8	0.24	0.28		

Table 1: Absolute relative differences to initial values of target publication type before and after transition

enough. Generating longer sequences of additional text, we find that the generated content is not connected and while we can find a chain of related topics of subsections, in some cases it is hard to define the common thread that is held throughout the whole text. Additionally, if such synonyms and paraphrases exist that corrupt the content but improve the relative differences to the targeted values of key text evaluation metrics, the methodology will replace existing words and sentences with corrupted content. Despite these drawbacks, we generated lots of results that reflect the targeted publication types and even more results that would require only minor changes to be completely acceptable. We conclude this article with satisfactory results of both content of generated texts and their values of key text evaluation metrics.

Our ideas for further work include improvement of natural language generation model, where the pre-trained model that we used should be trained on longer texts so that we could generate text based on longer prompts and thus make sure that we hold the common thread throughout the whole text. Determining whether synonyms or paraphrases corrupt the message of the text is also very important. Word embedding can be used to represent the context of the text and we could use it to determine whether the synonym fits the current context or not. Another way to adapt text to context would to create a dataset of texts, where each row hold different versions of the same text and each version represents the text written for different target audience. This way we would be able to teach text-to-text models to adapt text to context and it could also consider patterns that are not obvious to human's eye.

REFERENCES

[1] Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei, Elizabeth Trippe, Juan Gutierrez, and Krys Kochut. 2017. Text summarization techniques: a brief survey. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 8, (July 2017), 397–405. DOI: 10.14569/IJACSA.2017.081052.

[2] William B. Dolan and Chris Brockett. 2005. Automatically constructing a corpus of sentential paraphrases. In *Proceedings of the Third International Workshop on Paraphrasing (IWP2005)*, 9–16. https://www.aclweb.org/anthology/I05-5002.

[3] Ronen Feldman. 2013. Techniques and applications for sentiment analysis. *Commun. ACM*, 56, (April 2013), 82–89. DOI: 10.1145/2436256.2436274.

[4] Rudolf Flesch. 1979. *How to Write Plain English: A Book for Lawyers and Consumers*. Harper & Row.

[5] Kavita Ganesan, ChengXiang Zhai, and Jiawei Han. 2010. Opinosis: a graph based approach to abstractive summarization of highly redundant opinions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*. Coling 2010 Organizing Committee, Beijing, China, (August 2010), 340–348. https://www.aclweb.org/anthology/C10-1039.

[6] Cornelia Kiefer. 2019. Quality indicators for text data. In *BTW 2019 – Workshopband*. Holger Meyer, Norbert Ritter, Andreas Thor, Daniela Nicklas, Andreas Heuer, and Meike Klettke, editors. Gesellschaft für Informatik, Bonn, 145–154. DOI: 10.18420/btw2019-ws-15.

[7] Ibrahim Moawad and Mostafa Aref. 2012. Semantic graph reduction approach for abstractive text summarization. In (November 2012), 132–138. DOI: 10.1109/ICCES.2012.6408498.

[8] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2018. Language models are unsupervised multitask learners. https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf.

[9] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21, 140, 1–67. http://jmlr.org/papers/v21/20-074.html.

[10] Goutham Ramsri. 2020. Paraphrase any question with t5 (text-to-text transfer transformer). Online: https://towardsdatascience.com/paraphrase-any-question-with-t5-text-to-text-transfer-transformer-pretrained-model-and-cbb9e35f1555. *Towards Data Science*. . [Accessed: 17. 8. 2020]. (2020).