

ICMR 2019

8th International Conference on Multidisciplinary Research

WORDS OF PERSUASION: BEST ADJECTIVES FOR PERSUASIVE TED TALKS

Lim Zhong Heng (a), Gan Keng Hoon (b)*, Nur Hana Samsudin (c)

*Corresponding author

(a) School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia,
limzhongheng@student.usm.my

(b) School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia, khgan@usm.my

(c) School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia, nurhana.samsudin@usm.my

Abstract

As communication is a bridge connecting peoples in motivating changes, delivering ideas, influencing decision and create bonding; it is crucial to explore on the keywords that could help in persuading one's listener to accept the speaker ideas. The goal of this paper is to build a persuasive words framework that could identify most frequent adjectives used by the TED Talk speaker to make a public speaking to be persuasive. The proposed solution consists of four steps, namely Data Collection (Dataset Extraction), Data Preparation (Pre-process & Cleaning), Text Processing (Tokenize, Lemmatization, POS Tagging, Stop Word filter, Adjective Only Filter) and finally Extraction & Visualization of the words set. Case study on four different themes of talks: Technology, Global Issue, Science and Business were conducted based on the proposed model. Evaluation was carried out in two phases: internal validation and expert feedbacks were conducted to test the validity of the output. The output produced by the framework comprising of Word Cloud and Frequency Histogram that best describe each category of talk theme used in the evaluations. Lastly, discussions of each theme were carried out to further justify the findings and challenges of the generated output.

2357-1330 © 2020 Published by European Publisher.

Keywords: Text analytics, persuasive keywords, text extraction, ted talk.



1. Introduction

As public speaking is an important skill to have in either business or public relations arena, the skill of persuasion during the speak is the key to let the speaker to win over the crowd. The power of persuasion can carry everyone far in influencing decision, motivate changes and even forming strong bonds and connections for a person within the community. As it is commonly known that how persuasive a session of speech would be affected by tone, emotional influences and language fluency of the speaker, the persuasiveness of a speech is also closely depending on the context of the transcript. While there exist approach that leverage Deep Learning in analysing speakers' facial expression with their speech successiveness (Chollet, Wörtwein, Morency, & Scherer, 2016) and using advanced multimodal sensing as the assessment of speakers' public speaking ability (Fuyuno, Komiya, & Saitoh, 2018), there exist less supportive approaches that analyse the context of speaker's transcript in assessing their public speaking successiveness. In this paper, a series of text analysis processes are implemented to the famous speeches from well-known public speaking platform 'Ted Talk' to identify the most frequent adjectives used by the TED Talk speaker to make the speech to be persuasive.

2. Problem Statement

Communication is the backbone of our society. The quality of a communication could directly affect a person in forming connections, influencing decisions and motivate changes with their listeners. Public speaking is one the most important and most common form of communication that often needed either during a meeting, a report presentation or a speech giving session. The main purpose of public speaking is to deliver some important messages and motivate changes, as such, how effective the message to be delivered to the listeners is closely depends on the context of the transcript by the speaker. As such, the elements and characteristic of having a quality transcript is crucial to be identified in order to make the speech to be persuasive and able to bring up sympathetic response by the listener. According to studies (Ebaid, 2018), adjectives are defined as the integral elements in linguistic structures as it served as the main component to describe, identify and modify the noun, whereby the ubiquitous use of adjective would be a highly effective persuasive tool to a successive speech. Meanwhile, a recent work (Temple, 2018) has proposed a technique combining Natural Language Processing and Machine Learning to evaluate and identify the emotions of transcript's content and predicting the persuasiveness rating of 'Ted Talk' speech. The implementation of such methodology could identify the emotion created by the speaker to make their speech persuasive, however the author does not extract the influencing words that persuade the audience.

3. Research Questions

The study carried out in this paper will address: the following research questions.

- Can text processing pipeline be used to obtain the insights of Persuasive Words in the transcript of famous Ted Talks?
- Can the illustration and visualization of Persuasive Words usage help to identify the usefulness of such keywords among different Ted Talks categories?

4. Purpose of the Study

TED Talk served as a showcase for speakers in presenting great yet well-formed ideas in globally and tend to emerge more and more quality talks that are impressive and persuasive every year. These quality talks if analysed deeply, interesting insights could be extracted and served as the evaluation criteria in making a speech to be persuasive. Therefore, the conducted analysis is aimed to explore and identify on the commonly used adjective elements by the persuasive speech and less persuasive speech across different category of speech type using Text Analytics approaches.

5. Research Methods

Figure 1 shows the overview of the proposed persuasive words framework used by to analyse the insights of Ted Talks. The framework consists of four steps: Data Collection (Dataset Extraction), Data Preparation (Pre-Process & Cleaning), Text Processing (Tokenize, Lemmatization, POS Tagging, Stop Word filter, Adjective Only Filter) and finally Visualization of extracted word sets and elements. Each of these steps is described in the subsequent section.

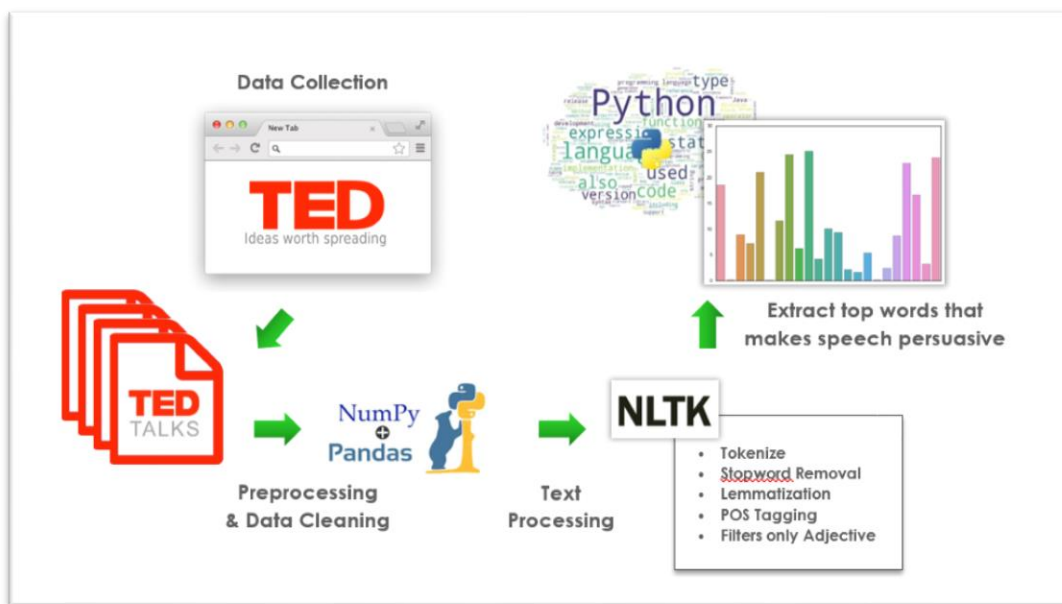


Figure 01. An overview of the persuasive words' framework for Persuasive & Non persuasive Ted Talk

5.1. Data Collection

The dataset is retrieved from Data World database by Owen Temple's Project, containing a complete listing of all TED Talks from official events posted in TED.com. The dataset includes the URL where video can be viewed, full English transcript URL, speaker name, headline, description, month and year filmed, event, duration, date published, topic tags and extended with abstract variables that describe the ratings of the talks, which is 'persuasive', 'inspiring' and 'unconvincing'. The rating variables are gathered by TED.com from their users who had voted for a particular talk using TED.com's ratings tool.

5.2. Data Pre-Processing and Data Cleaning

As the gathered dataset consist of too many irrelevant attributes, the first step of this section is to redefine a new dataframe with only relevant attributes: Transcript, Tag, Persuasive and ViewCounts. A complete data cleaning approaches (Fill NaN values, fix attribute types) is performed to make sure the data is cleaned.

Considering that the number of views over time since the talk posted might create a misleading data in 'Persuasive' attribute, a normalized attribute should be derived to better representing the persuasiveness of each talks. To normalize the rating to account for the number of views for each session, a new variable 'norm_persuasive' is extracted through dividing the count of persuasive votes by the number of times the talks had been viewed, which can be also defined as Persuasive votes per View of Talk (Figure 2).

transcript	# persuasive	# norm_persuasive
Good morning. How are you?()It's been	10704	226.6495
Thank you so much, Chris. And it's tr	268	83.7364
()Hello voice mail, my old friend.()I	230	140.5617
If you're here today ☐☐☐ and I'm ver	460	270.9788
About 10 years ago, I took on the tas	2542	211.7298

Figure 02. Derive Norm_Persuasive attributes

Meanwhile, each transcript is originally tagged with multiple topic category that best describe what the talks about. However, each title has multiple tags that used to describe the talks. To better segregate the speech type according to each tag allocated to each talk, each list of tags is first split into multiple rows attached with only single tag.

As the main objective of current work is to extract on the frequent words used in both persuasive and less persuasive talks, the instances are separated into persuasive (Norm_Persuasive \geq 250) and less persuasive (Norm_Persuasive $<$ 250) and categorized into top 4 tag category (Technology, Global Issue, Science and Business). The final dataset used is reduced to only the English transcript, tag, Normalized Persuasive Rating attributes.

5.3. Text Processing

After the cleaning process, each instance is tokenized using NLTK tokenizer (Bird, Klein, & Loper, 2009) into unigram bag of words. The reason of only applying unigram to the transcript is because we believed that a persuasive transcript might contain some 'Convincing Adjectives' that might better bring resonance across the listeners to the speaker. In such a way, unigram could be better in identifying the 'Adjectives' in POS tagging section at later.

During the tokenization of the transcript, a stop word filter is first applied to remove the commonly used word that does not contributes to later process. Each word is applied to retain only lower case before going into lemmatization process to further reduce the inflectional forms and derivationally related forms of a word to a common base form.

After applying lemmatization to the tokens, each word is classified into their parts of speech and labelling them accordingly to their lexical category using NLTK part-of-speech tagger. The motivations behind such steps is to identify the adjectives and retain only them for later uses. The tokens extracted to be only consisted of tags: 'JJ – Adjective', 'JJR – Comparative Adjective' and 'JJS – Superlative Adjective'. The frequency of each words is calculated using NLTK FreqDist to further validate the processed bag of words. Afterall, the same techniques and approaches are applied to both persuasive talks and less persuasive talks and ready to be further extraction of insights from the context.

5.4. Word Cloud and Frequency Histogram

After each word are processed and engineered to be fit into the wordcloud analysis, the cleaned data is loaded into Word Cloud and plotted with the frequency distribution histogram. The key steps in current section is to prepare a better and presentable visualization of data from the cleaned dataset, in another word is to prepare the bag of words illustrated informatic wordcloud and bar plot. Each set of instances (persuasive and less persuasive) is implemented with same methods to finally output their Word Cloud board and Frequent Word histogram. The most frequent words that contributes to a persuasive speech is then identified.

5.5. Evaluation Criteria

To further test and validate the suitability of the Word Cloud and Frequency Histogram, a set of formal evaluation and informal evaluation approach is conducted to evaluate the current model.

First session is to validate again the output of model (WordCloud and Frequency Histogram) through internal testing (Alpha test) on the retrieved bag of words. The challenge here is to recap the actual suitable words that appeared in the top 50 adjectives and fine tune the word tokens structure. The initial output is tending to have a lot of irrelevant nouns and adverb that particularly not making any sense. A process of choosing different part of speech tag in Penn Treebank tag set is iterated until the best word cloud that contains only adjective is generated.

Second session is to gather feedback from public speaking enthusiast and regular listeners using google doc survey form. The involved participants are explained with the motivations behind the project and required to comment and rate for the suitability of the WordCloud generated. Each participant is showed with different WordCloud according to different tag category (Technology, Global Issue, Science and Business) and let them rates on how helpful the output could be. The results show positive feedback experience as most participants give positive ratings to the output. Figure 03 shows the overview of expert feedback's results depicted in pie chart.

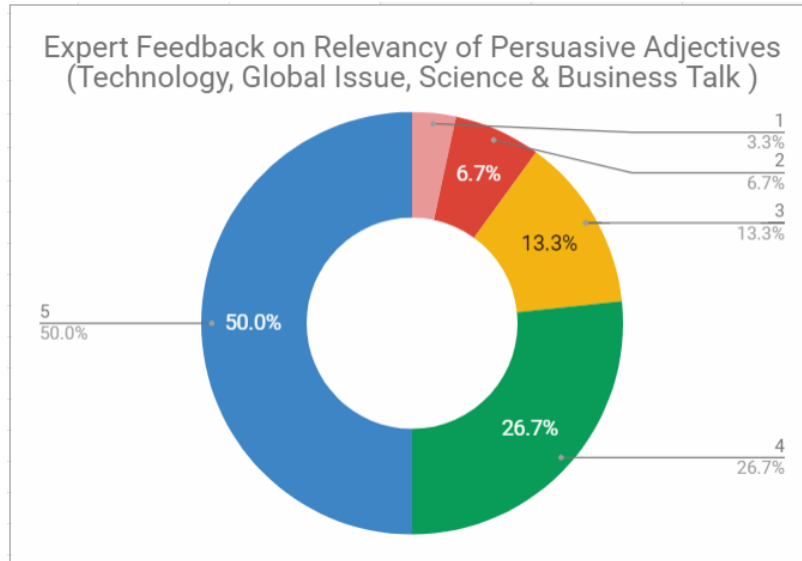


Figure 03. Derive Norm_Persuasive attributes

6. Findings

In this paper, four different themes of talks: Technology, Global Issue, Science and Business were analysed with proposed framework. The generated output is WordCloud and Histogram illustration of top 50 frequent words according to their category. The tokens used is well filtered to only contain adjectives that the speaker used during their talks that being rated as Persuasive by user of TED.com. Sections below depicts output of different categories and wrapped up with a discussions.

6.1. Technology Talk

The generated Word Cloud of Persuasive Technology is surprisingly containing more information than expected. As depicted in Figure 4, the top words used by the persuasive transcript is 'Antiangiogenic', 'Intergalactic', 'Cooperative', 'Testable', 'Recyclable'. The relevant topics that might be related to the words extracted is impressed, where adjective Anti angiogenic is used to describe the drugs that used in treatment to stop tumours from growing own blood vessels, Intergalactic is used to describe the travel between galaxies in science fiction and speculation, Recyclable is used to describe the waste that could be reuse or recycle. It is clearly describe that people are more convincing if the technology related stuff is related to these keywords' main purpose. While words like 'headtailhead', 'readwrite' could be categorized as irrelevant words as they do not provide much information.

For Less Persuasive transcript (Figure 5), there appeared to have some commonly used adjectives, such as Litte, many, first, different, much, good, great, able, big and so on. The insights here showcase that these adjectives are tended to be less persuasive when used in a technology talk.

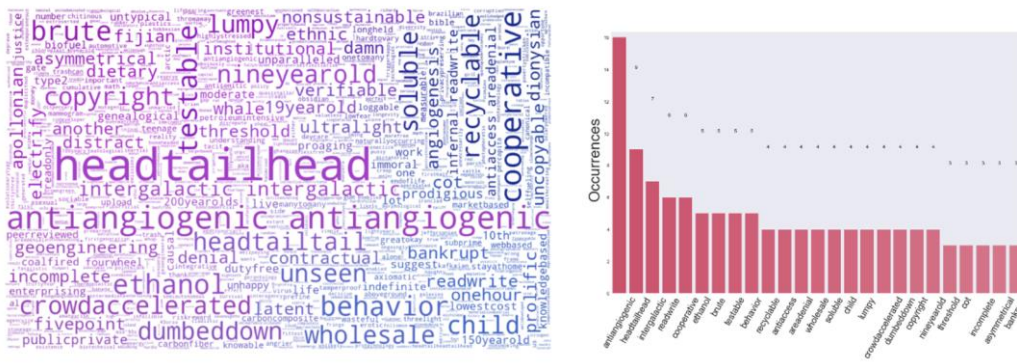


Figure 04. WordCloud & Frequency Histogram of Persuasive Technology Talk

For Less Persuasive transcript (Figure 5), there appeared to have some commonly used adjectives, such as Litte, many, first, different, much, good, great, able, big and so on. The insights here showcase that these adjectives are tended to be less persuasive when used in a technology talk.

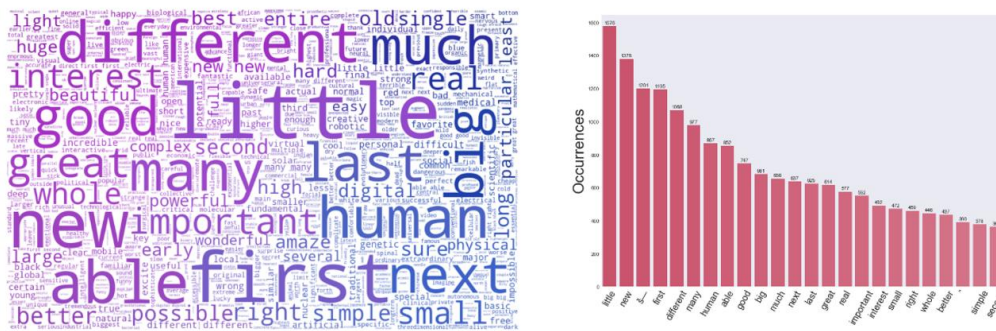


Figure 05. WordCloud & Frequency Histogram of Less Persuasive Technology Talk

6.2. Global Issues Talk

Meanwhile, the global issue talks transcript contained lesser information than expected and only certain relevant keywords is managed to be extracted. However, it is still visible that if adjective that related to ethical behaviour is being used, it has better chance to persuade the listener, for example Moral, Rational, Serious, Nonviolent and Liberal. Word Cloud of global issue talks transcript containing more irrelevant words (such as adverb) compared to technology talks, it is assumed that global issues topic might require more adverb style of adjective in supporting their emotional persuading techniques (Figure 6). While the less persuasive speech again showcases the commonly used adjectives just like technology, which can be assumed that the speech will be less persuasive if such keywords is being used during the talk (Figure 7).

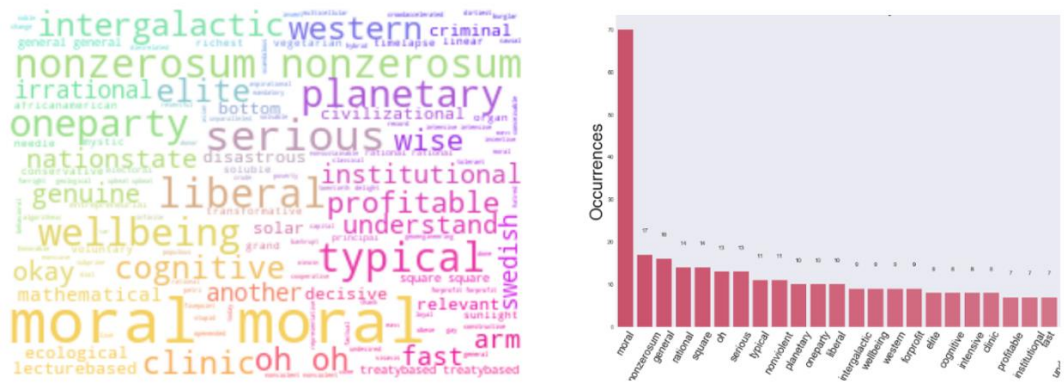


Figure 06. WordCloud & Frequency Histogram of Persuasive Global Issues Talk

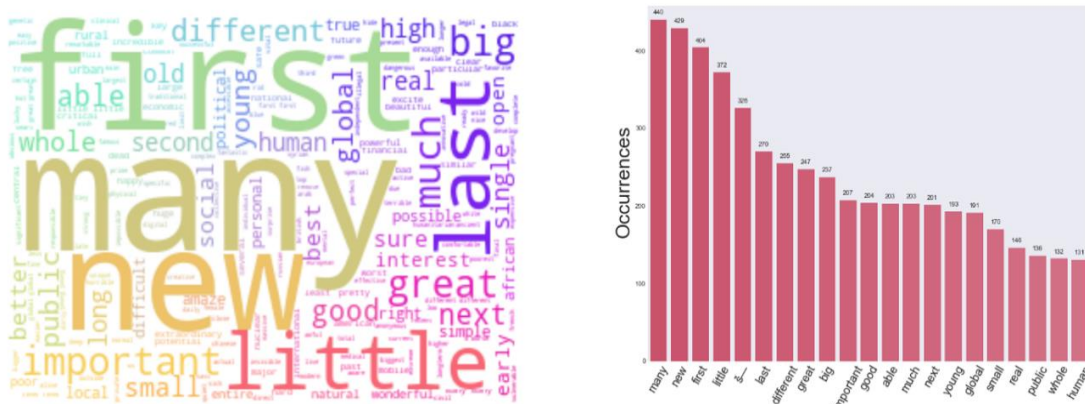


Figure 07. WordCloud & Frequency Histogram of Less Persuasive Global Issues Talk

6.3. Science Talk

For the science talk, the retrieved top adjective used in persuasive talk is surprisingly have some common words compared with Technology talk. Some adjective like Intergalactic, Antiangiogenic is again raised in the transcript, which again proof that these keywords could actually help in persuading their listeners and could be comprehended that these issues is more concerned by the listener. Key adjective that could raise resonance is Financial, Agnostic and Irrational (Figure 8) while the less persuasive transcript is again containing the commonly used adjectives (Figure 9).

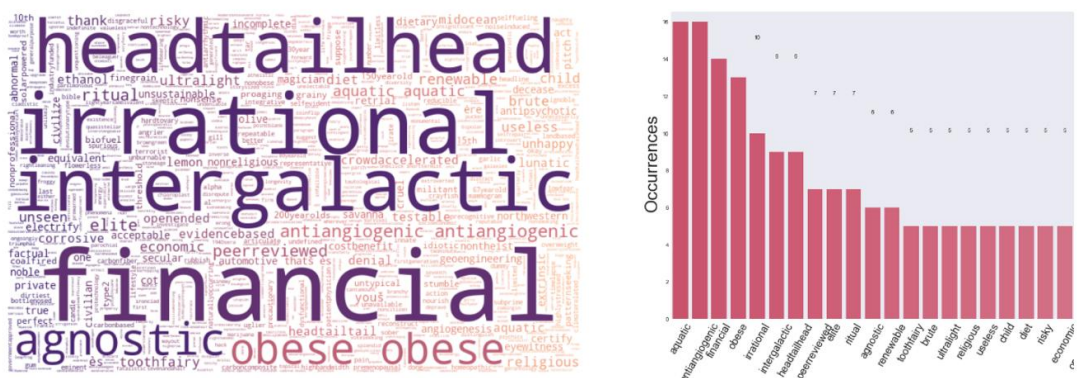


Figure 08. WordCloud & Frequency Histogram of Persuasive Science Talk

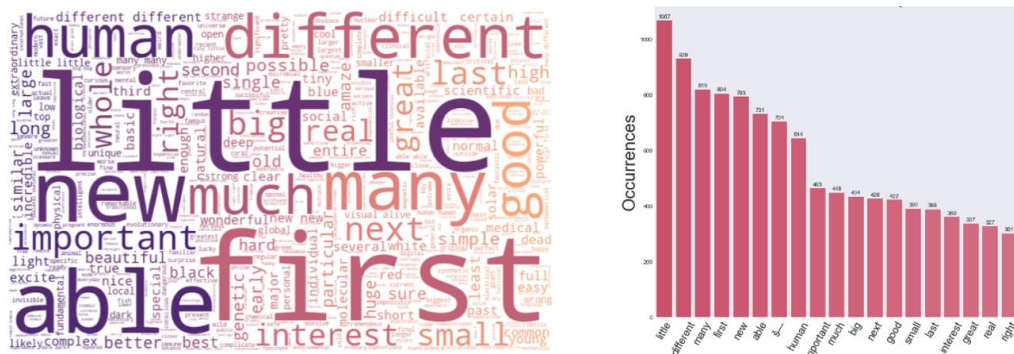


Figure 09. WordCloud & Frequency Histogram of Less Persuasive Science Talk

6.4. Business Talk

For business talks, it is surprisingly that the keywords extracted is more descriptive than being expected. The worth pinpoint extracted top words in persuasive transcript is varied across different categories, such as Moral, Copyright, Nuclear, Intangible, Tangible, Risky, Civil, Intrinsic and many more (Figure 10). Across these words, we could make assumption on a persuasive talk is tended to be rich in descriptive adjectives and more intense adjective. The less persuasive talks again contained the commonly used adjective that is less intense and does not makes them be convincing enough to persuade the listener (Figure 11).

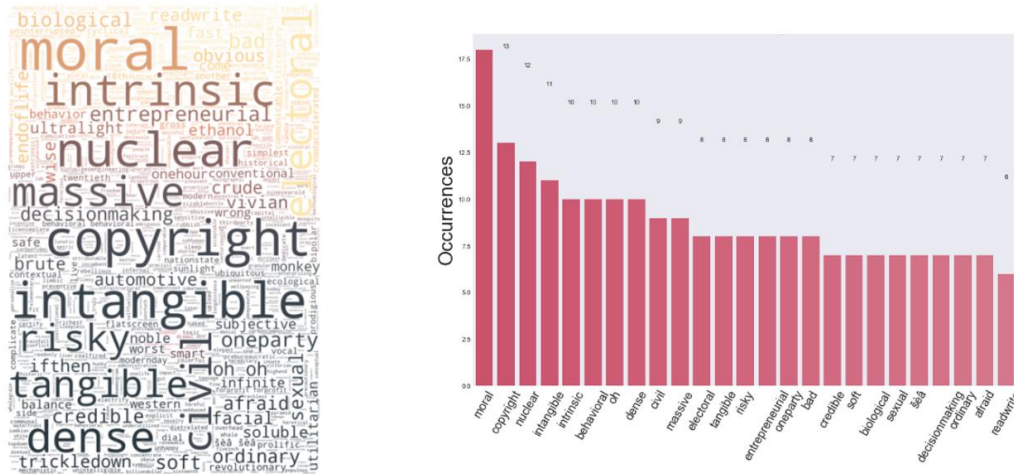


Figure 10. WordCloud & Frequency Histogram of Persuasive Business Talk

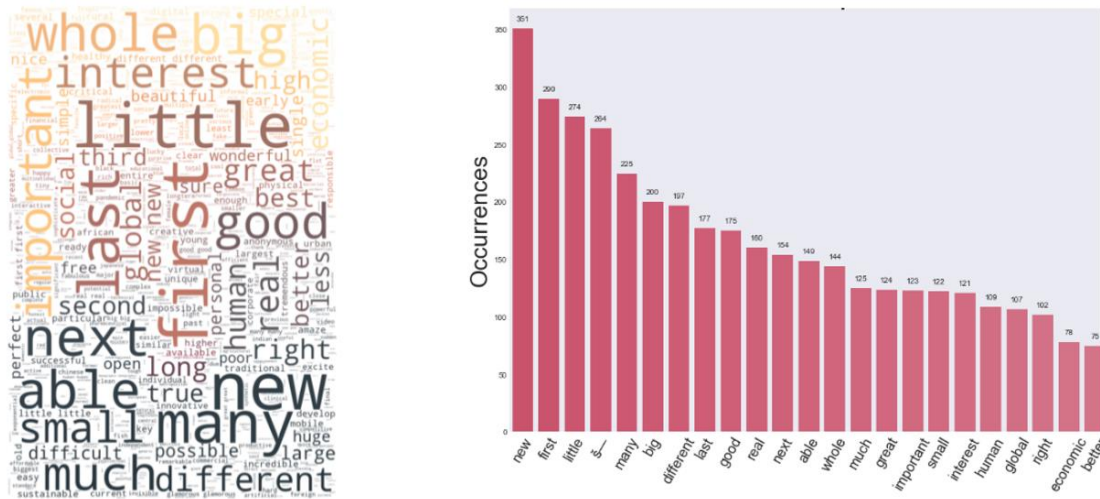


Figure 11. WordCloud & Frequency Histogram of Less Persuasive Business Talk

6.5. Discussions

From the work conducted, multiple text processing processes are performed to final generate a suitable model to better visualize on the frequent word used in persuasive speech and less persuasive speech. In the design stage, each process is compared with several similar techniques. Only suitable technique to was chosen in the design of the framework. In this section, we provide the discussions of each process and justify the selection of the selected techniques.

One challenge revealed during stop word filtering is the removal of punctuation of each word tokens. Although a list of stop words is used to filter the token list, but there still exist unwanted punctuation such as ‘100-year-old’, ‘low-power’ and more. Such a case is less efficient when plug into Word Cloud, where the word cloud module will separate the tokens into ‘100’, ‘year’, ‘old’ and provide an inaccurate output. To further fix this, the punctuation is manually removed with panda library function is left only word without ‘100yearold’, so each token is correctly used in word counting in Word Cloud module.

Besides, the tokens that will be filtered to contain only adjective might have different type adjective, such as superlative adjective and comparative adjective. To further reduce the inflectional forms and derivationally related forms of a word to a common base form, Lemmatization is applied. From there, the bag of words is then tagged with part of speech category and filtered to leave only the adjective for evaluation.

Lastly, as there appeared to have overlapped words set appeared in both persuasive and less persuasive transcript, we discovered that there is a need to segregate only the unique words that makes that particular speech to be persuasive. Considering the fact that some commonly used adjective would appear in both persuasive and less persuasive transcript, we compare both set bag of words and remove the duplicated word set, leaving only the unique word tokens to be our final data model for output generation.

7. Conclusion

As a conclusion, this work has successfully developed and applied the proposed framework on both persuasive talk and less persuasive talks based on different talk themes. From the results of extracted key

adjectives, we could conclude that for people are more convinced if the Technology talks contains the usage of adjectives in topics like *treatment for tumour*, *travelling between galaxies* and *recyclable products*. While for Global Issue talk, people are more convinced if the speaker stressed on emotional influencing adjectives. Listener in Science talk on the other hand is more persuadable if the speaker tends to link their speech to technologies describing adjectives. If the speaker wanted to persuade their listener better in a Business talk, they will have to include more intense and descriptive adjectives that could motivates and inspired the listener. Finally, for any topics of talk given, it is not recommended to only include the commonly used adjective, as these adjectives could lead to a less persuasive speech.

Acknowledgments

This research is supported by Fundamental Research Grant Scheme (FRGS), Ministry of Education Malaysia (MOE) under the project code, FRGS/1/2018/ICT02/USM/02/9 and title, Automated Big Data Annotation for Training Semi-Supervised Deep Learning Model in Sentiment Classification.

References

- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media Inc.
- Chollet, M., Wörtwein, T., Morency, L. P., & Scherer, S. (2016). A Multimodal Corpus for the Assessment of Public Speaking Ability and Anxiety. In *LREC*. Retrieved from <https://www.aclweb.org/anthology/L16-1078.pdf>
- Ebaid, H. A. (2018). Adjectives as Persuasive Tools: The Case of Product Naming. *Open Journal of Modern Linguistics*, 8, 262-293.
- Fuyuno, M., Komiya, R., & Saitoh, T. (2018). Multimodal analysis of public speaking performance by EFL learners: Applying deep learning to understanding how successful speakers use facial movement. *The Asian Journal of Applied Linguistics*, 5(1), 117-129.
- Temple, O. (2018). Words of Persuasion: Text predictors of Persuasive TED Talks. Retrieved from <https://data.world/owentemple/text-and-content-features-of-most-persuasive-ted-talks>