TED Talks by Phil Hansen and Vijay Kumar.

Sentiment analysis using R-programming.

Agunbiade, Adedoyin

Introduction

TED talks are influential public speeches delivered by experts from various fields. Phil Hansen, a multimedia artist, delivered a TED talk in 2013 titled "Embrace the Shake" where he discussed how he transformed his limitations into new possibilities by being creative, and he encourages others to do the same. His talk was inspired by his diagnosis of permanent nerve damage that prevented him from doing art. Vijay Kumar, a roboticist, delivered a talk in 2012 titled "Robots that Fly... and Cooperate." He discussed the challenges of building agile robots, the technologies involved, and their potential to revolutionise society. His robots can be used as first respondents or in construction, and their diversity and agility make them stand out. This report compares and analyses Phil Hansen and Vijay Kumar's topics, ideas, and contexts using text analysis tools in R programming.

Methods

Install Packages and Load Data: This analysis used R packages including `tidyverse`, `dsEssex`, `tidytext`, `ggrepel`, and `knitr`. The `ted_talk` dataset was loaded and filtered using the `filter()` function to include only talks by Phil Hansen and Vijay Kumar.

Pre-processing: To ensure consistency, the transcript data was cleaned by converting all characters to lowercase, removing unwanted characters, whitespaces, punctuations, and numbers. The unnest_tokens() function was used to tokenise the transcripts into individual words and convert data to tidy format thus enabling analysis with the packages. Stop-words were then removed using the `anti_join()` function on get_stopwords() to improve analysis efficiency and focus on meaningful words.

Word Analysis: Post-preprocessing, `tidyverse` package functions were used to determine word frequency and common words. Sentiment analysis was done via the `bing` and `nrc` lexicons to identify emotions in each speaker's speech. Log-odd ratios were calculated for each speaker's words using a custom compute_OR function, which includes a smoothing factor of 0.5 to address zero proportions. If the odds ratio >1, the first speaker is more likely to express the sentiment, while an odds ratio <1 suggests the second speaker is more likely to express the sentiment.

Data visualisation: `ggplot2` in `tidyverse` and `Knitr` were used to create barplot, scatterplot, and tables to help get better data insights and interpretation.

Result

Word Frequency analysis.

The initial word count between the speakers was 3886, which was reduced to 1877 after removing stop-words, indicating almost half of the words were irrelevant. Vijay Kumar used more words (1105) than Phil Hansen (772), suggesting Vijay Kumar used a greater number of vocabularies. This may be due to his longer talk duration and since this data did not consider the speech length, this result may be biased.

Figure 1 below shows the most common words used in Phil Hansen's and Vijay Kumar's TED talks.

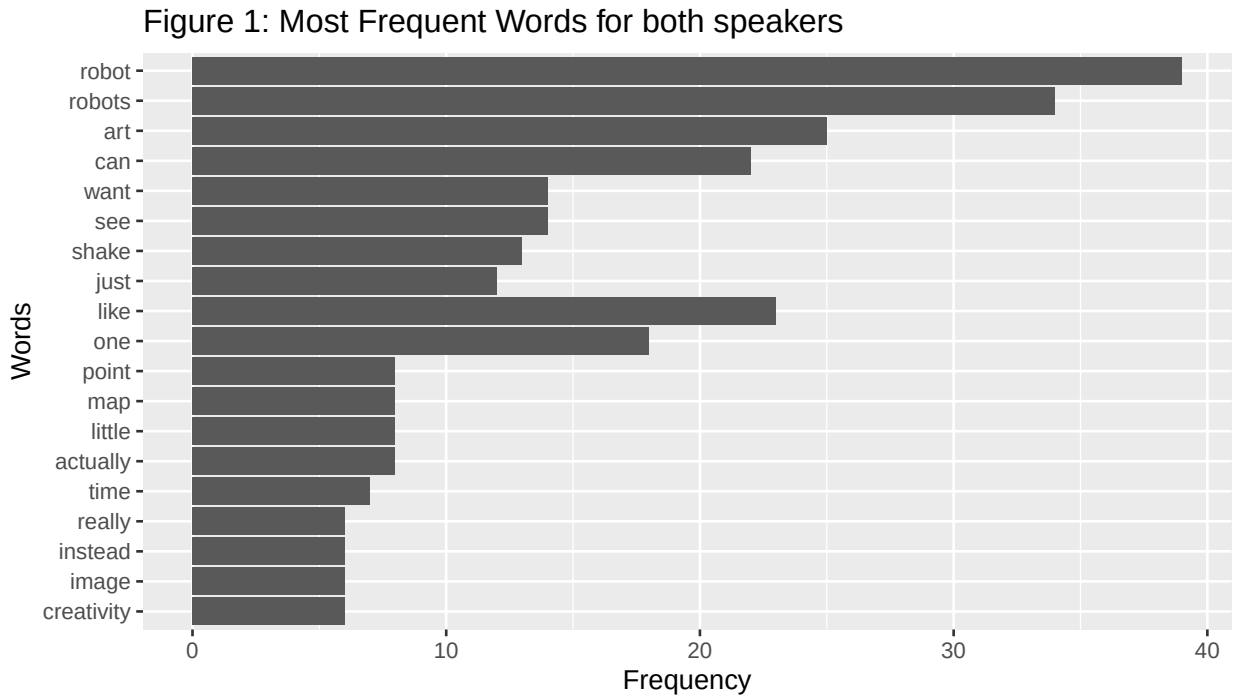
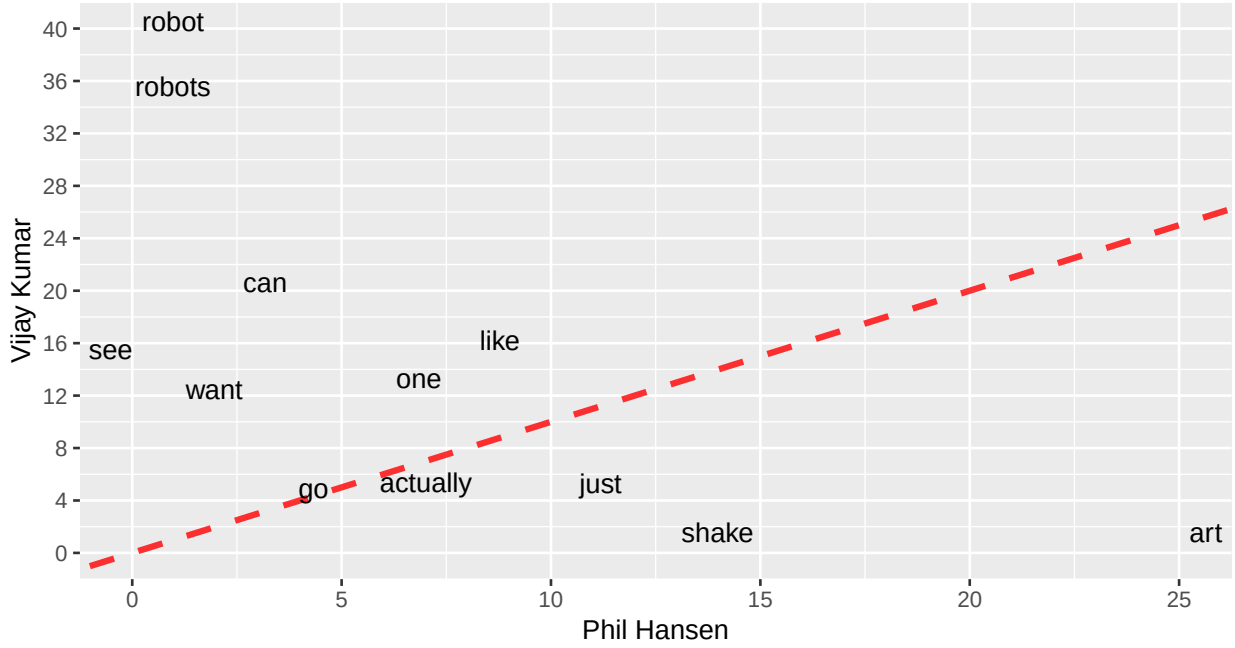


Figure 2 below also displays the word frequency analysis. The result shows that Vijay Kumar used the words "robot" (39 times), "robots" (34 times), and "can" (22 times) more frequently, while Phil Hansen used "art" (25 times), "shake" (13 times), and "just" (12 times) more often. The words "go" and "actually" were common between the two speakers, with "actually" being used more by Phil Hansen (8 times versus 7 times) and "go" was used more by Vijay Kumar (6 times versus 5 times). These findings indicate that both speakers have different approaches and perspectives on their topic, with Phil Hansen focusing on art and Vijay Kumar on robotics, and that their most frequently used words were related to their respective fields.

Figure 2: Comparison of words between Phil Hansen and Vijay Kumar



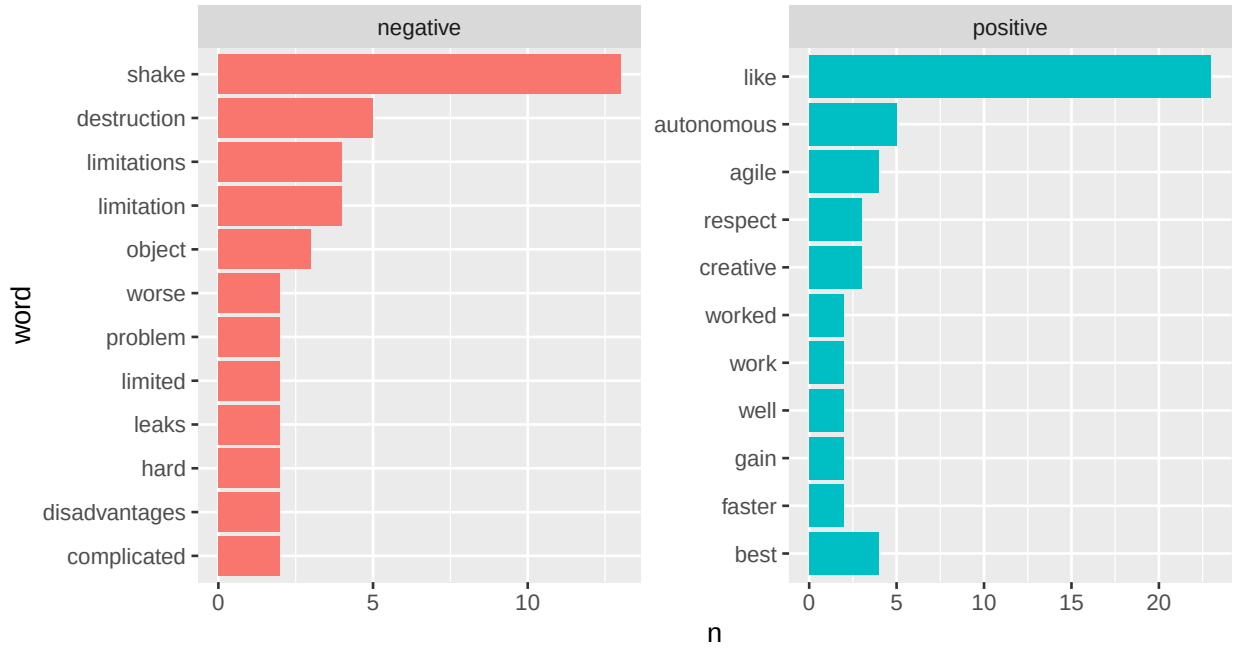
Sentiment analysis and log-odd ratio

Using the `bing` lexicon, it was displayed in Figure 3 that both speakers used both positive and negative sentiments. Some of the most commonly used negative sentiments were "shake" and "destruction while the most positively used sentiments were "like" and autonomous. Despite both speakers using positive and negative sentiments, their respective frequencies varied. Table 1 for instance demonstrates that Phil Hansen had a higher percentage of negative sentiments (63.8%) than positive sentiments (36.2%), while Vijay Kumar had a higher percentage of positive sentiments (68.7%) than negative sentiments (31.2%). When the speaker's sentiments were compared to each other, Vijay Kumar's talk had more positive sentiments (57) than Phil Hansen's talk (34), while Phil Hansen had a higher negative sentiment count (60) compared to Vijay Kumar's (26). Phil Hansen used words such as "creative" and "best" in a positive sense and words such as "shake" and "destruction" in a negative connotation. In contrast, Vijay Kumar used words such as "like" and "autonomous" positively and words such as "object" and "complicated" in a negative tone. Overall, these differences suggest that the speakers had different approaches to conveying their messages to the audience, which could influence how the audience perceives them.

Table 1: Table 1: Sentiment analysis and log-odd ratio of Phil Hansen and Vijay Kumar TED talk using BING lexicon

sentiment	Phil Hansen	Vijay Kumar	OR	log_OR
negative	60	26	3.8687783	1.352939
positive	34	57	0.2584795	-1.352939

Figure 3: Common sentiments for both speakers using bing lexicon



The log-odd ratio used to measure the odds of the sentiments conveyed helped in understanding the strength of association between a particular sentiment. The odd ratio (OR) and the log_odd ratio (log_OR) depicted in Table 1 above, support the analysis earlier described. The OR and log_OR for negative sentiment between Phil Hansen and Vijay Kumar were 3.86 and 1.35, respectively. This suggests that Phil Hansen is more likely to express negative emotions while Vijay Kumar is less likely to. Contrarily, the OR and log_OR for positive sentiments between the speakers were 0.25 and -1.35, respectively. Overall, the bing lexicon identifies Vijay Kumar as having a higher probability of expressing positive emotions, thus suggesting he is more optimistic, while Phil Hansen's words were more likely to express negative emotions suggesting his speech is more critical.

Furthermore, by using the `nrc` lexicon to identify the emotions conveyed, it was observed that the speakers used a range of sentiments, indicating that their talks were not entirely positive or negative. Table 2 result displays a slightly different result from Table 1's result. Analysing Table 2 and Figure 4 shows that even though Phil Hansen had a higher negative and positive sentiment count, these emotions are more associated with Vijay Kumar as the ORs were less than 1 and the log_ORs were negative. Table 2 result identifies that both speakers expressed anger and anticipation, but Phil Hansen exhibits these sentiments more. Also, the results demonstrate that both speakers equally expressed disgust (sentiment count of 8); however, this sentiment is more associated with Vijay Kumar, as shown in Figure 4. Figure 4 also shows that the most associated emotion for Vijay Kumar is "trust" while "surprise" was more associated with Phil Hansen.

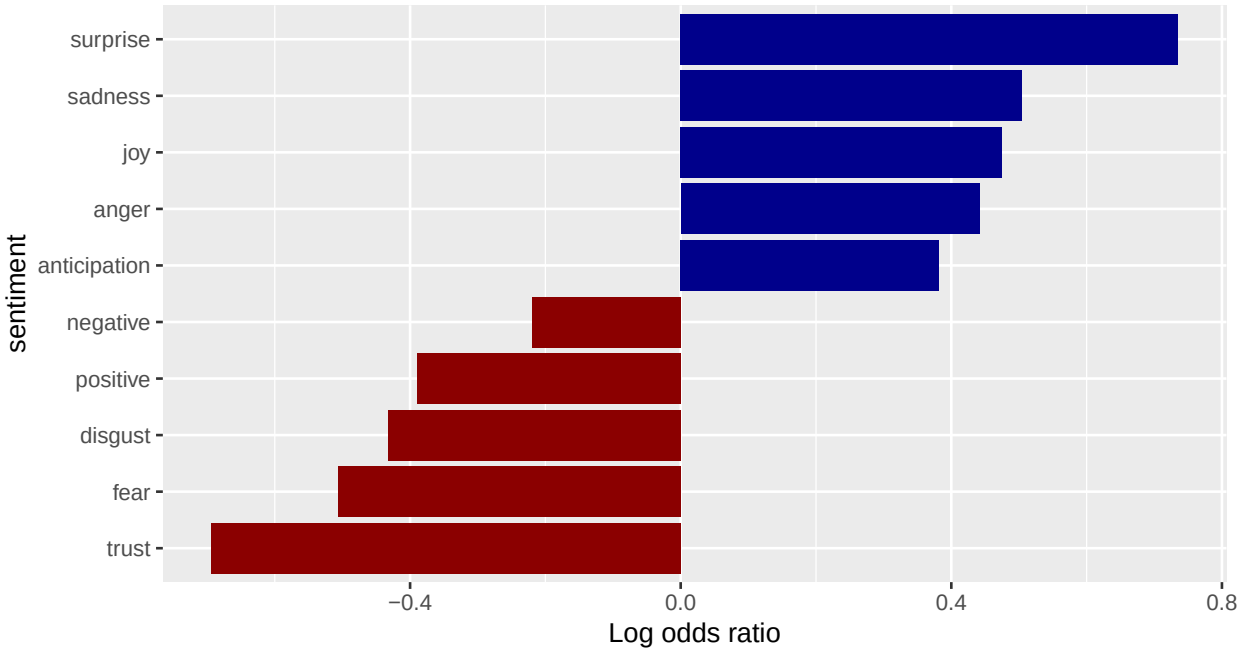
The two lexicons displaying slightly different analysis highlights the limitations of lexicons as its subjective and systematically different in word searches. Similarly, emotions conveyed may vary in a context that may not be reflected by lexicons. For instance, the word "trick" was used by Vijay positively; however, this was analysed as negative by the `bing` sentiment.

Another limitation of this analysis was that it was based solely on text and did not factor speaker's tone, body language, and visuals which could give better insight into the speaker's emotions and topic.

Table 2: Table 2: Common sentiments and log-odd ratio between Phil Hansen and Vijay Kumar using nrc

sentiment	Phil Hansen	Vijay Kumar	OR	log_OR
anger	21	9	1.5555556	0.4418328
anticipation	70	33	1.4649255	0.3818044
disgust	8	8	0.6485849	-0.4329624
fear	17	18	0.6030790	-0.5057071
joy	56	24	1.6072695	0.4745368
negative	35	28	0.8028967	-0.2195292
positive	91	80	0.6771628	-0.3898436
sadness	46	19	1.6558495	0.5043142
surprise	48	16	2.0859375	0.7352184
trust	40	48	0.4995748	-0.6939979

Figure 4: The association between sentiments using nrc lexicon



Conclusion

Overall, this text analysis helped gain insight into topics, ideas, context, and emotions involved in Phil Hansen's and Vijay Kumar's TED talk. Both speakers had different perspectives and unique speaking styles, and their words related to their fields. Different emotions were used to convey their message; however, more analysis and other factors need to be considered to gain a better understanding.