Distributed Sentiment Analysis: Geopolitical News Coverage

Anastasiia Golovchenko et al.

ABSTRACT

This study aims to design and implement a scalable system architecture for large-scale extraction and sentiment analysis of YouTube video transcripts using advanced speech-to-text technologies and distributed computing with Apache Spark. The proposed methodology is demonstrated through a comparative sentiment analysis of news coverage on the Israel-Iran conflict across different news broadcasters. The system integrates the Google Cloud Speech-to-Text API for transcription and TextBlob for sentiment analysis, with data extraction, preprocessing stages, and sentiment analysis parallelized using Apache Spark. The proposed speech-to-text system achieved an accuracy of 88.74% for news videos with a single reporter, surpassing the performance of the Google English ASR model. The obtained sentiment scores and trends aligned with the actual events, demonstrating the correctness and precision of the methodology. The study highlights the influence of editorial stance, target audience, and regional perspectives. The potential development of a real-time 'news monitor' based on the proposed methodology seems viable, providing valuable insights for decision-makers, researchers, and the general public.

CCS CONCEPTS

• **Computing methodologies → Distributed computing methodologies**; **Speech recognition**; **Lexical semantics**; **Machine learning approaches**.

KEYWORDS

YouTube, Sentiment Analysis, Apache Spark, Distributed Computing, Speech-to-Text, Big Data, Video Transcripts, Scalability

1 INTRODUCTION

YouTube, as a platform for user-generated content, has become an invaluable source of data for researchers and analysts. The platform's diverse video content and user base offer a unique opportunity to gain insights into public sentiment on various topics. However, the volume of data and the complexity of processing video content pose significant challenges in extracting and analyzing this information. Therefore, leveraging advanced speech-to-text technologies and distributed computing for the large-scale extraction and sentiment analysis of YouTube video transcripts can be a solution to address these challenges and unlock novel insights from this rich repository of opinionated data.

While traditional text data sources, such as news articles and social media posts, have been widely studied, YouTube video transcripts remain a relatively untapped resource for sentiment analysis. The primary objective of this study is to design and implement a scalable system architecture capable of processing vast amounts of video data, applying sentiment analysis techniques to the extracted transcripts using Apache Spark, and laying the groundwork for future large-scale analysis, which can be applied to a wide range of additional purposes beyond sentiment analysis.

The specific goal of this study lies in the potential for comparative sentiment analysis about the Israel-Iran conflict across different news broadcasters, such as BBC News (UK), CNN (USA), SABC News (South Africa) and Al Jazeera English (Qatar). Applying the proposed techniques to news videos from these sources provides a deeper understanding of how sentiments regarding this specific topic vary across different regions and media outlets, demonstrating the practical utility of the applied approach as a task-specific value-added analysis.

Looking ahead, the proposed system can be extended to develop a 'news monitor' that automatically processes batches of audio data from multiple news providers and ranks sentiments on different topics in real-time. This application highlights the importance of the scalability achieved through distributed computing, as it enables the efficient processing of continuous streams of data from diverse sources. The potential impact of this study extends across various domains, including market research, public opinion mining, and social science research, by introducing a new, rich source of data, providing a scalable methodology for processing it, and offering valuable insights into public sentiment at scale.

2 BACKGROUND AND RELATED WORK

Sentiment analysis, a subfield of natural language processing (NLP), has gained significant attention in recent years due to its wide-ranging applications in understanding public opinion, market trends, and social dynamics [11]. Traditional sentiment analysis techniques have primarily focused on textual data sources, such as social media posts[17], product reviews [18], and news articles [4]. However, the growth of video-sharing platforms like YouTube has opened up new opportunities for sentiment analysis research.

Early research on sentiment analysis of YouTube content primarily focused on analyzing text-based comments and metadata. Khan et al. investigated YouTube video comments, assessing the video's rating by analyzing user feedback and revealing the importance of user sentiments using the SentiStrength tool [9]. Sushma et al. reviewed sentiment analysis methods for YouTube videos, extracting text comments and transcripts and categorizing them into neutral, positive, negative, relevant, and irrelevant along with a transcript summary levels [22]. Melnyk and Feld explored various methods for detecting the stance of German YouTube comments on gender diversity, compared the results with sentiment analysis, and found that these are distinct NLP tasks focusing on different discourse characteristics [14]. Including sentiment parameters into the training data was used for examining comments on YouTube videos in the study by Baravkar et al. It was suggested that using genuine comments obtained from the YouTube API can be employed to train the model for more accurate decision-making [1].

As the field progressed, more advanced natural language processing techniques and machine learning algorithms were applied to enhance the accuracy and scope of sentiment analysis on YouTube. Mehta et al investigated sentiment analysis of YouTube ad views using various machine learning algorithms, including Linear Regression, Support Vector Machine, Decision Tree, Random Forest, and Artificial Neural Network. They conducted a comparative analysis based on the experimental results obtained from these models to determine the most effective approach for predicting sentiments in YouTube ad views [13]. Poria et al. proposed a multimodal sentiment analysis framework that combines visual, audio, and textual features to predict sentiments in YouTube videos. Their approach demonstrates the importance of considering multiple modalities for accurate sentiment prediction [19]. Similarly, Zadeh et al. introduced the CMU-MOSI dataset, a multimodal sentiment analysis dataset consisting of YouTube videos, which has become a benchmark for multimodal sentiment analysis tasks [23].

While these studies highlight the potential of sentiment analysis in video content, they often focus on small-scale datasets and do not address the challenges of processing large volumes of data. To tackle this issue, researchers have turned to distributed computing frameworks like Apache Spark. Sirisha et al. proposed using Hadoop technologies like HDFS and Hive to store and analyze massive video data from YouTube [21]. Jalil et al. utilized the memory computation framework 'Apache Spark' to extract social media data and perform sentiment analysis on both textual and visual data, aiming to provide a method for better analysis of specific data [8].

However, there is a notable gap in the literature regarding the large-scale extraction and sentiment analysis of YouTube video transcripts using distributed computing frameworks. The current study aims to bridge this gap by proposing a YouTube video analyzer that integrates sentiment analysis. Furthermore, the current study focuses on a specific use case of analyzing sentiments related to the ongoing geopolitical tensions between Israel and Iran in order to gauge public opinion on a specific subject across a diverse range of media sources. The proposed system's scalability and potential for real-time analysis align with the growing interest in sentiment analysis for news and blogs [2].

3 METHODOLOGY

3.1 Data Extraction and Collection

In this study, a data extraction and collection approach was employed to obtain text transcripts from YouTube videos. The process began by leveraging the capabilities of the Google YouTube Data API v3 [7], which allows for searching and retrieving video metadata based on specific criteria, such as channel ID, video duration, and query terms.

To demonstrate the effectiveness of the methodology, a specific use case was implemented focusing on extracting video URLs from the BBC, CNN, SABC News and Al Jazeera English YouTube channels, specifically related to the Israel-Palestine conflict. By utilizing a Python client library for the YouTube Data API, the study first identified the channel IDs of the targeted news organizations through search queries. Subsequently, these channel IDs were used to refine the search for videos pertaining to the chosen topic and date range with additional filters applied to limit the results to short-duration videos. The extraction was limited to short-length videos for the last 30 days to optimize the data collection process and ensure more focused and relevant content. [1]

Having obtained the video URLs from the API response, the study proceeded to download the audio content of each video using the pytube library, a lightweight, dependency-free Python library designed for this purpose [20]. This streamlined approach enabled the efficient retrieval of audio files without the need to store the entire video data.

To ensure compatibility with the Google Cloud Speech-to-Text API, which was employed for transcription, the downloaded audio files underwent a conversion and adjustment process. This involved transforming the audio format to WAV, modifying the sample rate to 16000 Hz, and setting the number of channels to mono.

The adjusted audio files were then uploaded to a Google Cloud Storage (GCS) bucket [5], which serves as a scalable and reliable storage solution.

Utilizing Apache Spark's Resilient Distributed Datasets (RDDs), the study efficiently managed the parallel extraction of video URLs and downloading of audio files. This method not only accelerated data handling but also enhanced fault tolerance and resource optimization across the distributed computing environment.

3.1.1 Performance evaluation of parallelization of audio download and upload. To assess potential performance gains and demonstrate scalability, the processing time for audio download, standardizing and uploading the audio from 15 BBC news videos were analyzed. The videos were of varying length, ranging from 47 seconds to 15 minutes, with most around 6-9 minutes. The experiment was conducted on Google Cloud Platform (GCP) using the standard cluster setup from week 9 of the course, with the number of workers adjusted between 2 and 3. The performance was evaluated across different number of RDD partitions. For comparison, a benchmark test using a Pandas data frame completed the same process in 96 seconds.[2]

Figure 1 shows that using PySpark even with just 1 partition is significantly faster than using pandas. This can be attributed to optimized in-memory processing, efficient management of overheads, and advanced optimizations, for example for I/O operations.

Generally, increasing the number of partitions tends to decrease the processing time for both worker configurations (2 and 3 workers). This is expected as more partitions allow for better load distribution across the cluster, leading to more efficient use of the available computational resources. The most significant drop in processing time happens when going from 1 to 2 partitions, and increasing the partitions consistently improves performance until a certain point, beyond which the returns diminish. This happens around 6 partitions in our test. Too many partitions can lead to increased overhead in terms of scheduling and task management. If each partition's task duration is very short compared to the overhead of managing the partition, then having a large number of partitions can actually degrade performance.

With 3 workers, not only does the processing time generally lower across all partition settings compared to 2 workers, but the reduction in time is also more pronounced with fewer partitions.

[1] Code for **URLs extraction**.
[2] Code for **Performance evaluation of parallelized setup for audio download and upload**.

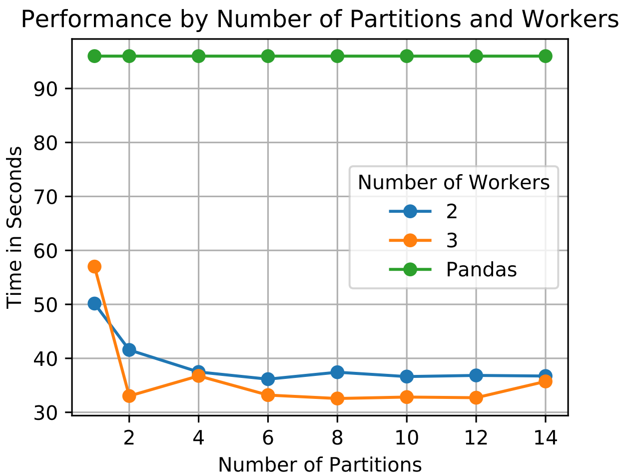

Figure 1: Time taken to download, standardise (convert to one channel and 16.000 MHz), and upload audio to GCS bucket for 15 videos of varying length.

For instance, with 2 partitions, the time drops significantly from around 41 seconds (2 workers) to 33 seconds (3 workers). The performance increase when going from 2 to 3 workers is likely to be even more significant as the number of videos is scaled. However, the performance gain should be compared to the additional cost of the increasing computational resources.

Since our task involves downloading and uploading audio files, scaling would probably lead to the operation being I/O bound rather than compute-bound. Therefore, the optimal number of partitions and workers would depend not just on CPU and memory but also on the network bandwidth and the I/O capabilities of the storage system. Too many concurrent uploads or downloads might saturate the network bandwidth, negating the benefits of increased parallelism. Monitoring not only the execution time but also system metrics like network utilization could aid our understanding of this problem.

Furthermore, the variation in video lengths could impact the outcomes of this test. When certain videos are substantially longer than others, the nodes that end up handling this partition may become bottlenecks. While this may be noticeable in our small test dataset of 15 videos, as the dataset expands and each partition increases in size, the distribution of video lengths across different partitions should become more uniform.

Note that the variability in results, such as the observed spike time spike with 4 partitions for 3 workers, might be caused by fluctuations in activity across the GCP during the test. To get more consistent data, it is recommended to perform multiple iterations of the test at various times of the day and compute the average of the outcomes.

3.2 Data Processing and Conversion

Once the audio data is available in the GCS bucket, the study employs the Google Cloud Speech-to-Text API, a highly accurate and scalable solution for converting audio data into text transcripts. This API utilizes advanced machine learning models to accurately transcribe speech, even in the presence of background noise or multiple speakers. [3]

Although ideally suited for parallel processing, we encountered challenges in implementing this across multiple nodes on GCP due to dependency management issues with Python packages. This limitation prevented us from fully leveraging parallel computing to speed up the transcription process. Resolving these issues could have significantly enhanced our data retrieval speeds. Additionally, the process of transcribing speech to text was repeatedly interrupted due to age verification requirements for sensitive content, and attempts to resolve the issue by uploading an ID and confirming age were unsuccessful.

To verify the data quality and robustness of the sentiment analysis, the study considers additional text processing techniques. One approach involves measuring the accuracy of the transcription process by comparing the generated transcripts with a set of manually transcribed audio files. This accuracy measurement follows the industry-standard Word Error Rate (WER) metric, which calculates the percentage of incorrect word transcriptions in the entire set. The WER is computed by aligning the machine-generated transcript with the human-provided ground truth transcript and counting the number of substitution, deletion, and insertion errors [6].

Basic pre-processing steps are applied to ensure data quality, including converting the text to lowercase, removing punctuation marks, removing multiple spaces, and stripping leading and trailing whitespace. The evaluation was conducted across three categories of videos: news with multiple reporters, news with a single reporter, and conversational-style videos (TED talks), using the Jiwer library in Python for WER calculation [3]. [4]

Measuring the WER enables the assessment of the Google Cloud Speech-to-Text API's accuracy for the specific domain of videos and helps identify potential areas of improvement and ensures the reliability of the generated text data for subsequent sentiment analysis.

3.3 Dataset

The resulting dataset is a consists of video metadata and processed text transcripts extracted from a range of YouTube videos across different news channels, specifically focusing on the Israel-Iran conflict in April.

The included features are:

- Metadata extracted from the YouTube API (video title, URL, channel, video duration, date of publishing): This approach enables filtering videos for various experimental setups, such as focusing on specific time periods or themes by applying filters to the titles. In cases where performance is suboptimal, it is also feasible to exclude longer videos by filtering on video duration.
- Processed text transcript: The Speech-to-Text API includes an option to include timestamp for the transcript which could be used for another experimental setup.

[3]Code for **Audio extraction and Speech-to-Text transcription**.
[4]Code for **Accuracy checking**.

Including these features enhances the flexibility and adaptability of the setup, allowing for easy adjustment to accommodate various experimental designs and to scale the study by, for example, extending the time period under consideration.

3.4 Sentiment Analysis and Scalable Machine Learning

With the processed text data available, the study proceeds to perform sentiment analysis using scalable machine learning techniques. Sentiment analysis aims to determine the overall sentiment expressed in each text transcript, providing insights into the emotional tone and opinions conveyed in the videos.

Although building a custom model was considered, the extensive task of gathering data to create an effective and robust sentiment model, which may not have definitively yielded improved results, led to the decision to utilize a prebuilt model. This study employs a prebuilt model, TextBlob, which is a freely available Python library that provides a simple API for various natural language processing tasks, including sentiment analysis [12]. The sentiment analysis model in TextBlob is trained on movie review data from the NLTK corpus using a Naïve Bayes model. It outputs sentiment scores as a range of values, with polarity scores between -1.0 (negative) and 1.0 (positive) and subjectivity scores between 0.0 (very objective) and 1.0 (very subjective).

The text transcripts obtained from the Speech-to-Text API undergo a preprocessing step for sentiment analysis. The preprocessing involves lowercasing the text for uniformity and removing punctuation that does not contribute to sentiment, while keeping exclamation marks as they can intensify sentiment. The preprocessed text is then passed to the TextBlob library, which internally handles tokenization and uses raw text tokens for training, eliminating the need for lemmatization. The sentiment polarity and subjectivity scores are calculated for each text transcript. [5]

Leveraging TextBlob's model efficiently processes the extracted video transcripts and assigns sentiment scores to each transcript. The scalability of this approach is ensured by integrating the sentiment analysis process with the distributed computing capabilities of Apache Spark, enabling parallel processing of large volumes of text data. [6]

First of all, a sentiment analysis is conducted on BBC news videos covering the Israeli-Palestinian conflict over the past 30 days. The goal is to identify trends in the polarity scores and understand how the sentiment fluctuates in response to significant events during this period. Furthermore, the sentiment of the conflict-related videos is compared to the overall sentiment of all BBC news videos during the same timeframe in order to determine whether the polarity of the conflict coverage deviates from the general sentiment expressed in BBC news content. The analysis then explores the sentiment trends of BBC news videos covering other topics, such as China and Trump, over the last 30 days to compare sentiment trends between different events. Finally, the sentiment of BBC news videos covering the conflict is compared to the sentiment of Al Jazeera news videos on the same subject over the 30-day period. This comparison seeks to identify any notable differences between the two news sources, potentially revealing variations in their coverage of the tensions.

In addition, word clouds were generated to visualize the most frequently occurring words in the positive and negative sentiment texts for a set of videos from BBC, CNN, and SABC News over the last 30 days. The NLTK library is used to obtain a set of English stopwords, which are further expanded to include words specific to the Israel-Palestine conflict. The video transcripts are classified as 'positive' or 'negative' based on their polarity scores, and the word clouds are created using the WordCloud library, with the specified stopwords removed [15, 16][7].

4 RESULTS

4.1 Accuracy Evaluation

The accuracy of the automated speech-to-text transcriptions was evaluated and compared across three categories of videos: news with multiple reporters, news with a single reporter, and informal TED speeches. The average WER for the three groups was 19.80% for news with multiple reporters, 11.26% for news with a single reporter, and 22.36% for TED speeches, respectively (Figure 2). The results indicate that the news videos with a single reporter generally had the lowest WER, suggesting that the system performed best on this category, achieving an accuracy of 88.74%. This could be attributed to factors such as clearer audio quality, more consistent speech patterns, and fewer overlapping speakers in these videos. In contrast, TED speeches had the highest average WER, indicating that the system faced more challenges in accurately transcribing this category, with an accuracy of 77.64%. This may be due to the more diverse and complex nature of the content, including the use of specialized vocabulary, varied speaking styles, and the presence of audience reactions.

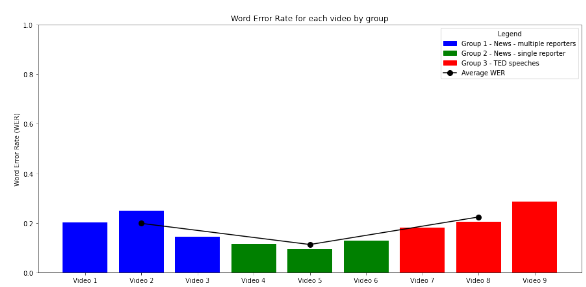

Figure 2: Word Error Rate for videos by group

4.2 Polarity Score Over Time

The polarity score from BBC news videos related to the Israel-Palestine conflict over the last 30 days is presented in Figure 3.

[5]Code for **Sentiment analysis**.
[6]Code for **Parallelization of sentiment analysis**.
[7]Code for **Generating word clouds**.

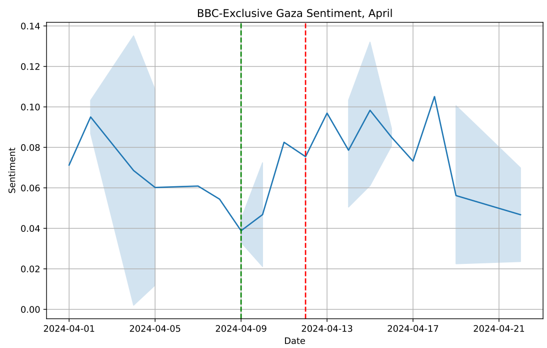

Figure 3: BBC News Polarity Score: Israel-Palestine Conflict, April 2024

The polarity score starts relatively high on 2024-04-01 but sharply declines on 2024-04-05, likely due to Iranian rallies and promises of retaliation after Israeli strikes on Iran's consulate. The polarity score reaches its lowest point on 2024-04-09, possibly attributed to a Gaza aid delivery incident, which resulted in the deaths of over 100 people. A slight rebound in polarity is noted, but it falls on the 13th, related to Iran's first direct attack on Israeli territory, marking a significant escalation in the conflict. On April 19th, Israel launched a missile against Iran in response to Iran's action, further intensifying the tensions between the two nations. The polarity score remains low towards the end of the period, likely influenced by the potential for further escalation between Israel and Iran.

To gain further insights into the sentiment trends, the sentiment of the conflict-related videos is compared to the overall sentiment of all BBC news videos during the same period (Figure 4).

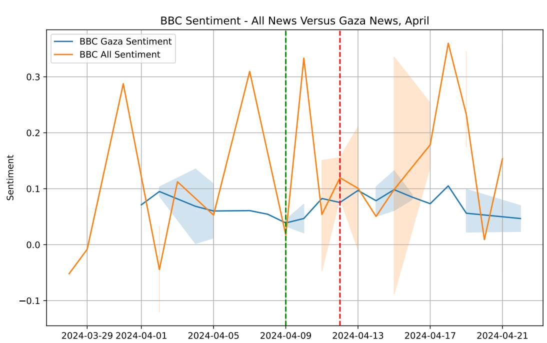

Figure 4: Comparative Sentiment Trends: Israel-Palestine Coverage and all BBC News, April 2024

The sentiment of the videos related to the Israeli-Palestinian conflict has less pronounced fluctuations and a consistently more negative tone compared to the overall sentiment of BBC news. The overall BBC news sentiment, which covers a wide range of topics, is more likely to experience higher fluctuations compared to the sentiment of a single topic like the Israeli-Palestinian conflict. This is because the broader scope of news coverage encompasses various events across different categories, which might be significantly positive or significantly negative, each with its own sentiment implications. In contrast, the coverage of a specific topic, such as the Israeli-Palestinian conflict, tends to be more focused and consistent in its sentiment, as the news stories are centered around a particular issue and its related events.

To better understand the previous sentiment analysis, it is essential to explore the sentiment patterns of coverage on other specific topics to gain a more comprehensive understanding of how different events are portrayed.

As illustrated in Figure 5, the sentiment scores for China-related news exhibit the most pronounced fluctuations among the analyzed topics, suggesting that the news coverage of China encompasses a wide range of favorable and unfavorable events, similar to the diverse nature of all BBC news sentiment. Trump-related news shows a generally negative stable trend with a couple of significant fluctuations, likely reflecting his recent lawsuits and criminal cases. In general, the conflict-related sentiment has a more pessimistic score with numerous fluctuations, although not as considerable as those of China and Trump. This reflects the gravity and ongoing nature of the events.

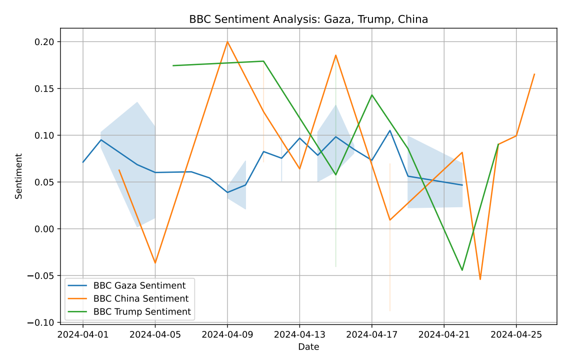

Figure 5: Comparative Sentiment Trends: Coverage of Israel-Palestine Conflict, China, and Trump by BBC News, April 2024

The comparative sentiment analysis of BBC and Al Jazeera's coverage of the Israeli-Palestinian conflict reveals some notable differences, despite the overall negative sentiment trend throughout the period (Figure 6). In the early part of the month, Al Jazeera's sentiment scores are slightly more negative compared to BBC's, suggesting that Al Jazeera's coverage may have been more critical or focused on the negative aspects of the conflict during this time. Interestingly, the Israeli attack on a humanitarian aid convoy on 2024-04-09 appears to have had a more pronounced impact on Al Jazeera's sentiment score compared to BBC's, with Al Jazeera's score dropping more sharply.

However, in the latter half of the month, after Iran's first direct attack on Israel, BBC's sentiment scores become more negative than Al Jazeera's. This could be explained by the fact that BBC, being a UK-based news organization, may be more aligned with Israel's position in the conflict, potentially leading to a more negative sentiment in their coverage following an attack on Israeli territory. The sharp rise in Al Jazeera's score at the end of April might be attributed to the news that Hamas received an Israeli proposal amid efforts to revive Gaza talks.

Therefore, while analyzing different news channels, it's important to consider factors such as editorial stance, target audience, and regional perspectives that may influence the sentiment of their coverage. The comparison between BBC and Al Jazeera sentiment scores highlights that even when covering the same conflict, news sources may exhibit variations in their sentiment due to differences in their approach, focus, and interpretation of events.

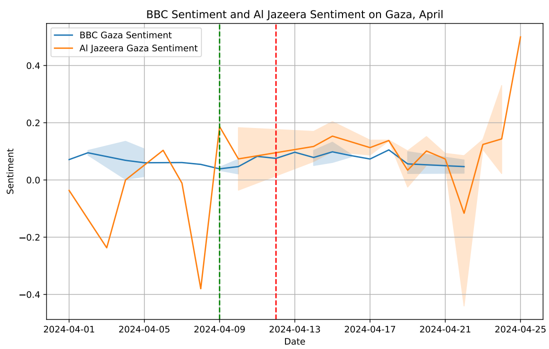

Figure 6: Comparative Sentiment Trends: Israel-Palestine Coverage by BBC News and Al Jazeera English, April 2024

4.3 The Word Cloud Visualizations

The word cloud visualizations provide insights into the overall sentiment and frequently used words in the video transcripts related to the Israel-Palestine conflict. Figure 7 displays the word cloud for negative sentiment texts, highlighting words such as 'strike', 'war', 'fighting', 'conflict', 'crime', and 'dangerous', which are associated with the negative aspects of the Israel-Palestine conflict.

Figure 7: Word cloud for negative sentiment texts

Figure 8 presents the word cloud for positive sentiment texts, featuring words like 'peace', 'time', 'saying', 'well', 'kind', 'like', and 'know', indicating a more optimistic and hopeful perspective on the situation.

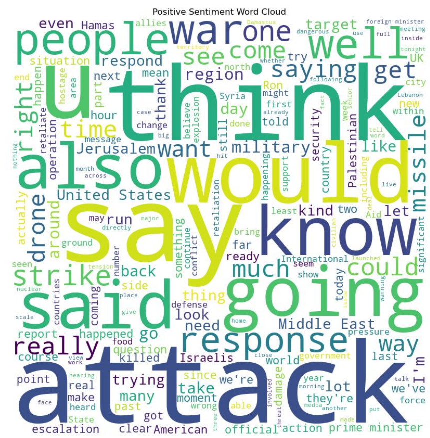

Figure 8: Word cloud for positive sentiment texts

Interestingly, some words such as 'people', 'war', and 'want' appear in both the positive and negative sentiment word clouds. This overlap can be attributed to the context in which these words are used. For instance, 'people' may be mentioned in both positive and negative contexts, such as discussing the impact of the conflict on people's lives or highlighting people's desire for peace. The presence of these words in both word clouds underscores the complexity of the Israel-Palestine conflict.

5 CONCLUSION

The scalable system architecture proposed in this study demonstrates the feasibility of processing large volumes of video data for sentiment analysis by integrating the Google Cloud Speech-to-Text API with Apache Spark for distributed computing. However, it is important to note that while the data extraction, preprocessing stages and sentiment analysis were successfully parallelized using Apache Spark, the transcription process itself could not be fully parallelized due to technical issues encountered with GCP. Specifically, challenges related to dependency management for Python packages on multiple nodes prevented the complete parallelization of the speech-to-text conversion. Despite this limitation, the overall system architecture still showcases the potential for distributed computing to handle large-scale video data processing and sentiment analysis. Future work could focus on resolving the technical hurdles associated with parallelizing the transcription process on GCP or exploring alternative cloud computing platforms like Microsoft Azure or Amazon Web Services (AWS) to fully leverage the benefits of distributed processing for this critical step in the workflow.

The performance evaluation of the parallelized audio download and upload processes using Apache Spark demonstrates significant performance gains compared to using a Pandas data frame. Increasing the number of partitions and workers generally led to decreased processing times, with the most substantial improvements observed when transitioning from one to two partitions. However, the study also emphasizes the importance of considering factors such as network bandwidth and I/O capabilities when scaling the system. Future research should explore the optimal balance between the number of partitions, workers, and available computational resources to maximize performance while minimizing costs.

The proposed speech-to-text system achieved a high level of accuracy of 88.74% for news videos with a single reporter, which surpasses the performance of the Google English ASR model reported by Kuhn et al. [10], with a WER of 20.1% in 2024. This demonstrates the effectiveness of the proposed system in accurately transcribing news content and its potential for various applications such as content indexing, subtitling, and accessibility. However, it is important to note that the accuracy varied depending on the video category, with ordinary life speeches presenting the greatest challenges. Future work should focus on refining the system to handle more diverse video categories and further improving its accuracy and robustness.

The obtained sentiment scores and trends can be explained by the actual events, demonstrating the correctness and precision of the methodology. However, it is essential to consider that editorial stance, target audience, and regional perspectives can influence the sentiment expressed by different news sources, even when covering the same events. This underscores the importance of considering multiple viewpoints and sources to gain a more comprehensive understanding of complex geopolitical issues. Future research could explore the use of a wider scale of sentiment scores, moving beyond the standard range of -1 to 1, to capture more nuanced variations in sentiment. Additionally, sentiment scores could be adapted to specific regional characteristics, taking into account cultural, linguistic, and sociopolitical factors that may affect the interpretation and expression of sentiment in different contexts. This approach would enable a more fine-grained and context-sensitive analysis of sentiment trends across diverse regions and news sources.

The overlap of certain words in both the positive and negative sentiment word clouds highlights the nuanced nature of the conflict, emphasizing the need for sentiment analysis approaches that can capture and interpret the contextual use of language.

Furthermore, the potential development of a real-time 'news monitor' that automatically processes audio data from multiple sources and ranks sentiments on different topics seems viable and underscores the practical utility and scalability of the proposed methodology. Such an application could provide valuable, up-to-date insights for decision-makers, researchers, and the general public, enabling them to stay informed about shifts in public opinion and sentiment across various issues and regions.

REFERENCES

[1] Aditya Baravkar, Rishabh Jaiswal, and Jayesh Chhoriya. 2020. Sentimental Analysis of YouTube Videos. *International Research Journal of Engineering and Technology (IRJET)* 07, 12 (Dec 2020).

[2] Mikhail Bautin, Lohith Vijayarenu, and Steven Skiena. 2008. International Sentiment Analysis for News and Blogs. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 2. 19–26.

[3] Digital Darwin. 2013. Jiwer: Simple CLI app with a collection of commonly used tasks. https://github.com/digitaldarwin/jiver.

[4] Xiaowen Ding, Bing Liu, and Philip S. Yu. 2008. A Holistic Lexicon-Based Approach to Opinion Mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*. 231–240.

[5] Google Cloud Storage [n. d.]. Google Cloud Storage Documentation. https://cloud.google.com/storage/docs. Accessed: 2023-04-17.

[6] Google Speech-to-Text [n. d.]. Measure and improve speech accuracy. https://cloud.google.com/speech-to-text/docs/speech-accuracy. Accessed: 2023-04-17.

[7] Google YouTube API [n. d.]. Google YouTube Data API v3 Documentation.

[8] Shaikh Ayesha Jalil, Sawant Namrata Vikas, Madye Sonal Madhusudan, and Prof. Pravin Jangid. 2020. Social Media Sentiment Analysis Using Big data For textual and visual data. In *Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 1–6.

[9] Mohammed Arsalan Khan, Sumit Baraskar, Anshul Garg, Shineyu Khanna, and Asha M. Pawar. 2021. YouTube Comment Analyzer. *International Journal of Scientific Research in Computer Science and Engineering* (Aug 2021).

[10] Korbinian Kuhn, Verena Kersken, Benedikt Reuter, Niklas Egger, and Gottfried Zimmermann. 2024. Measuring the Accuracy of Automatic Speech Recognition Solutions. *ACM Transactions on Accessible Computing* 16, 4 (2024), 1–23.

[11] Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies, Vol. 5. Morgan Claypool Publishers. 1–167 pages. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

[12] Steven Loria. 2022. TextBlob: Simplified Text Processing. https://textblob.readthedocs.io/. Accessed: 2023-04-27.

[13] Tanvi Mehta and Ganesh Deshmukh. 2022. YouTube Ad View Sentiment Analysis Using Deep Learning and Machine Learning. *arXiv.org* (2022).

[14] Lidiia Melnyk and Linda Feld. 2023. On Methods of Data Standardization of German Social Media Comments. *Journal of Computer-Assisted Linguistic Research* 7 (2023), 22. https://doi.org/10.4995/jclr.2023.19907

[15] Andreas Mueller. 2023. wordcloud 1.9.3 - A little word cloud generator in Python. https://pypi.org/project/wordcloud/.

[16] NLTK Project. 2023. Natural Language Toolkit (NLTK). https://www.nltk.org/.

[17] Alexander Pak and Patrick Paroubek. 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. 1320–1326.

[18] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*. 79–86.

[19] Soujanya Poria, Erik Cambria, Newton Howard, Guang-Bin Huang, and Amir Hussain. 2016. Fusing Audio, Visual and Textual Clues for Sentiment Analysis from Multimodal Content. *Neurocomputing* 174 (2016), 50–59.

[20] pytube [n. d.]. pytube: A lightweight, dependency-free Python library for downloading YouTube videos.

[21] Y Sirisha, K. S. Parimala, and Arun K Jyothi. 2017. You Tube Data Analysis Using Hadoop Technologies Hive. In *Proceedings of International Conference on "Smart and Sustainable Initiatives in Natural Sciences and Engineering" 2017*, Vol. 1. Malaysia, 10–16.

[22] Goli Sushma, Vadicherla Raju, Rajashekar Kandakatla, Neethipudi Sashi Prabha, Rajesh Saturi, and L. Mohan. 2024. YouTube Video Analyzer Using Sentiment Analysis. *International Journal of Intelligent Systems and Applications in Engineering* 12, 16s (2024), 597–601. https://ijisae.org/index.php/IJISAE/article/view/4897

[23] Amir Zadeh, Rowan Zellers, Eli Pincus, and Louis-Philippe Morency. 2016. MOSI: Multimodal Corpus of Sentiment Intensity and Subjectivity Analysis in Online Opinion Videos. *arXiv preprint arXiv:1606.06259* (2016).