O'Halloran, K. L., Pal, G. & Jin, M.H. (2021). Multimodal Approach to Analysing Big Social and News Media Data. *Discourse, Context and Media* (40) https://doi.org/10.1016/j.dcm.2021.100467

Multimodal Approach to Analysing Big Social and News Media Data

Kay L. O'Halloran[a], Gautam Pal[a] and Minhao Jin[b]

[a]Department of Communication and Media, University of Liverpool, Liverpool L69 7ZX, United Kingdom

[b]Department of Computer Science, University of Liverpool, Liverpool L69 3BX, United Kingdom

Abstract

Multimodal analysis traditionally involves conceptualising abstract frameworks for language, images, and other resources and their intersemiotic relations (e.g. text and image relations) and then demonstrating these frameworks with some examples. This scenario has changed with the recent move towards multimodal approaches to big data analytics which will involve empirically testing and validating multimodal theory and frameworks through the analysis of large data sets. However, large training sets of analysed texts are required to develop computational models based on multimodal theory. Therefore, an alternative approach which involves integrating multimodal frameworks with existing computational models for big data, cloud computing, natural language processing, image processing, video processing, and contextual metadata is proposed. The integration of these disparate fields has the potential to dramatically improve computational tools and techniques, thus placing multimodality at the forefront of research aimed at mapping and understanding multimodal communication. As a step forward in this direction, we explore how existing computational tools and approaches can be integrated into a multimodal analysis platform (MAP) with facilities for searching, storing and analysing text, images and videos in online media, together with dashboards for visualising the results. Preliminary analyses and classifications of text and images about COVID-19 and George Floyd in five online newspapers and Twitter postings show how media patterns can be studied using existing computational tools. The study highlights (a) the benefits and current limitations of big data approach to multimodal discourse analysis and (b) the need to incorporate knowledge about language, images, metadata, and other resources as semiotic systems (rather simply sets of symbols and pixels) to improve computational techniques for big data analytics.

Keywords: Multimodal analysis; news media; social media; Twitter; COVID-19; George Floyd

1. Introduction

Interest in multimodal analysis has surged over the past few decades, as evidenced by publications in this area (see overview in O'Halloran, 2020; Tan et al., 2020). The traditional approach adopted within multimodal studies is to conceptualise abstract frameworks for language, images, and other resources and their intersemiotic relations (e.g. text and image relations) and demonstrate these frameworks with some examples. Recent efforts have shifted towards empirically testing and validating multimodal frameworks through the analysis of large data sets (e.g. O'Halloran et al., 2018; O'Halloran et al., 2021). However, one major problem is that large training sets of analysed texts are required in order to develop computational approaches based on multimodal theory. This obstacle is difficult to overcome with limited manpower and resources. Therefore, an alternative approach is proposed which involves integrating multimodal frameworks with existing computational models for big data, cloud computing, natural language processing, image processing, video processing, and contextual metadata. In addition to validating multimodal frameworks, the integration of these disparate fields has the potential to improve computational tools and techniques, thus placing multimodality at the forefront of research aimed at mapping and understanding multimodal communication.

As a step forward in this direction, we demonstrate how existing computational tools for natural language processing and image processing can be integrated into a multimodal analysis platform (MAP) to analyse large datasets of online mainstream newspaper articles and social media posts, in this case, Twitter. This is not a simple undertaking, however, as will be seen in the ensuing discussion. First, a common platform with processing pipelines needs to be developed to collect, store, and analyse the different media and display the results. This is a complex task, given that data needs to be secured from various online media sources and organised in a consistent format so it can be processed and analysed. Second, the approach raises fundamental questions about how text and image data can be aggregated and analysed computationally. That is, snapshots of various analyses undertaken with language and image processing tools can be displayed using dashboards. However, these standard analyses offer limited insights on their own, particularly as natural language processing (NLP) tools largely operate at the rank of word or word group with the result that grammatical and discourse patterns remain largely unexplored. Most importantly, the issue remains of how to integrate the text and image results. As will be seen in this study, one approach involves developing linguistic descriptors of the visual images so that NLP algorithms can be applied to the language and image data. In turn, this data can be converted to numerical values (in the form of vectors) for further analysis. However, this raises questions in relation to the changes and reduction of meaning which occurs when visual images are converted to linguistic descriptors. Furthermore, issues such as assigning weights to the results of the linguistic and visual analyses and capturing and aggregating these results across language and image data remain largely unresolved. Importantly, MAP is fully extendable for integrating with broad range of analytical techniques, making it suitable for further technological and theoretical advances beyond what is presented here. As such, MAP provides a productive environment for both conducting multimodal analysis *and* learning about multimodality, creating the opportunity to link theory and practice across multiple disciplinary domains (e.g. Bateman et. al, 2017). We return to this issue in the final section of this paper.

2. Computational Approaches to Multimodal Analysis

Software models, despite recent advancements in machine learning and big data technologies, significantly lack sophisticated methods for analysing the aggregation of text, image, and video data (e.g. Seng et al., 2019). Consequently, the state of the art computer models, in reality, are still monomodal in nature – that is, the models are individually designed for language processing, image and video processing and audio processing. In addition, collecting, storing, and analysing large datasets of online media in a single integrated multimodal analysis platform remains a major logistical challenge. That is, divergent technologies are required to crawl the web in real-time to extract a large volume of text, image, and video data quickly and accurately and index these materials to make the data immediately-searchable for the creation of interactive analytical reports. The data collection process should be capable of collecting both historical and real-time datasets based on a date range and search criteria for meaningful insights on online news and social media analysis. However, current approaches of collecting online media within a limited time range cannot authentically represent data as it contains a large amount of noise and unrelated records. For example, to develop comprehensive insights on topics such as COVID-19 and George Floyd's death[1], multiple sub-topics need to be aggregated together along with processing and interaction with different modalities like text, image, video, historical batch data, and real-time streams. The data needs to be harvested from multiple channels simultaneously encompassing both social and news media, and noise must be filtered out to improve the accuracy of the results.

According to Social Media Monitoring Tools and Services Report published by Ideya Ltd (2018), there are at least 157 social media analysis tools are available, with the majority being paid subscriptions.

None of the tools, however, provides an analysis of both news and social media data within the same application, given the complexity of analysing different genres and their various deployment of language, images and other semiotic resources (e.g. Bateman, 2017). Existing tools can be used to analyse the social media platforms including Twitter, Facebook, Instagram, YouTube, Reddit, Pinterest, LinkedIn, and Tumblr, but these same tools cannot aggregate the data across the platforms (Maynard et al., 2013). Moreover, while these tools focus on language analysis, more social media tools are employing artificial intelligence (AI) and image recognition to tackle visual content (e.g. see Ideya Ltd, 2018). At this stage, social media tools do not analyse aggregated textual and image analysis data across social media platforms. The technological innovation of combining language analysis *and* visual analysis (including video content) across media platforms lies at the heart of the MAP project. Also, tools like Sprout Social[2] can be used to analyse Twitter data to evaluate hashtag performance, track Tweet clicks and measure Facebook page impact, and LinkedIn connections evaluate Pinterest post sentiments, for example. However, the majority of the tools lack advanced filtering options and thus cannot be used for complex queries (for example, "Tweets that are related to Donald Trump and the coronavirus and contain both images and videos posted from Washington"). At the current time, commercially available tools offer easy to use web-based interfaces that allow users to graphically interact with the system without the need to write software code (e.g. Atrey et al., 2010; Dumas et al., 2009). Applications, therefore, can easily be used by social scientists without any knowledge of data science or programming languages. Nevertheless, in-depth analysis often requires custom functionalities that extend beyond the out of the box features that exist within these tools. For this reason, advanced analytics tools need to incorporate the custom data science logic through Structured Query Language (SQL) or programming languages like Python and Java.

Sophisticated image analysis, such as reverse image search, is offered by existing platforms (e.g. Guinness et al., 2018; Reilly and Thompson, 2017; Tan et al., 2018). Visual evidence of the number of times an image is used across social media platforms is a powerful metric, but the same frameworks cannot aggregate multiple image searches together. Moreover, images, videos, and textual results are interpreted separately instead of producing a joint multimodal analysis produced by the platform. In this regard, a comprehensive and overall outcome is thus only achieved through a manual process of interpretation from each of the reports separately.

Furthermore, traditional multimodal analysis is based on data stored in flat-files or Relational Database Management System (RDBMS) as a single standalone desktop (Hoel et al., 2005; Larios et al., 2012). These tools cannot offer an effective solution for a collaborative, centralized application where multiple users are able to simultaneously access the system. Computing processing power needs to be scaled up incrementally with a fresh inflow of real-time stream data that are added up with existing historical data. Without indexing the large volume of textual data, applications fail to achieve a useful interactive reporting experience as each report can take several hours to complete. In summary, current applications fall short of combining multiple modalities in regard to enhancing the objective of multimodal understanding.

Based on the aforementioned review, in summary, MAP exploits the following limitations with the commercially available tools:

- **Functionality.** The existing platforms we examined only offers textual data processing, although technological advances are starting to target visual content (e.g. Ideya Ltd 2018).
- **Performance.** Without using indexing and data caching mechanism, the search performance degrades.
- **Convenience**. Need to use different set of tools for collecting social and news media data.
- **Cost.** Apart from subscriptions fees, data collection cost is significant through APIs. For example, collecting 50,000 Tweets costs in excess of $100.

In what follows, we describe the multimodal analysis platform (MAP) which was developed for the optimal fusion of disparate modalities (text, image and videos) from multiple data source streams. MAP minimizes the manual interpretations of the combined modalities in a way that is capable of: (a) efficient information extraction, ingestion, and data cleansing; (b) effective integration of data from different media sources in a cross-media interaction scheme; (c) leveraging the latest technologies in data science, big data, NLP and image processing; and (d) easy to use intuitive web-based interface, making technological advancements in multimodal analysis accessible to the non-data scientists.

The facilities in MAP are demonstrated through a case study involving the analysis of online newspaper articles from five UK online newspapers and Twitter posts about COVID-19 and George Floyd, an African-American man killed by police during an arrest in Minneapolis in the United States on 26 May 2020 (British Summer Time). The analysis is undertaking during three time periods: (a) the week before George Floyd's death; (b) the day George Floyd died, and (c) the week following his death. The aim is to analyse the language and images in these articles and social media postings in order to track discourse trends during this pivotal time in history. The variety of tools which are integrated into MAP for this case study are described below.

3. Features of the MAP Platform

The cloud-based multimodal analysis platform (MAP) for real-time data collection, big data analytics, and reporting was developed as a solution to the limitations of the existing tools for multimodal analysis. Significantly, MAP permits various tools to be integrated and tested, leading to a greater understanding of what can be achieved with existing tools and what new computational functionalities need to be developed to address current limitations. At this stage of development, the three primary facilities in MAP are:

1. *Social and News Media Data Collection*: The platform permits text, image, and video data to be collected at scale, with accompanying metadata. At the current time, data is being collected from Twitter and five leading UK newspapers *The Guardian*, *The Independent*, *The Evening Standard*, *The Metro,* and *The Sun*. Data from news media is collected at scheduled intervals by an automated process from a list of newspapers. For Twitter, the user explicitly collects data through a real-time Twitter search using a keyword or user profile name.
2. *Indexing and Semantic Annotation*: The Data Collection process generates a number of JSON formatted files that are subsequently indexed to improve search performance. Image and video data are semantically annotated with linguistic labels using the image and video analysis tools provided by Clarifai[3]. These tools include general models for identifying objects, themes, moods, and demographics of persons in the image (for age, gender, and cultural appearance).
3. *Search and Interactive Reports*: Multiple search results are combined at the data collection step through a web-based interface. The aggregation is applied across Twitter and multiple newspapers for text, image, and video by clicking on the appropriate options in the interface. Any unwanted data is filtered using the advanced Structured Query Language (SQL) filter, date range, data source type (e.g. text, image or both) and stop words. The user provides SQL queries and stop words in textboxes, and the remaining filters are applied using a graphical interface: for example, radio buttons, checkboxes and so forth. The SQL filters can be used to exclude or include dataset based on any field values (e.g. filtering Twitter hashtags related to COVID which are retweeted more than 1 million times since February 2020). The resulting charts and tables can be exported in PDF, CSV, JSON, or XML formats.

These facilities are considered in turn below.

3.1 Social and News Media Data Generation

MAP is currently hosted on Google Cloud Platform[4], which offers a range of cloud computing services that run on the same infrastructure which Google provides for its own products (e.g. Google search, Gmail, file storage and so forth). MAP is hosted on a Linux instance of the Google Compute Engine[5], the Infrastructure as a Service (IaaS) component of Google Cloud Platform. As shown in Figure 1, the raw data stream is first loaded into the flat files and in the downstream processing pipeline, data moves to the Splunk[6] indexer after data cleansing by using stop words to remove irrelevant parts of the data for this case study (e.g. newspaper navigation links such as “search jobs”, “sign in”, search buttons, and content such as advertisements, weather reports, etc). Splunk software is designed for capturing, indexing, and correlating real-time data in a searchable repository from which graphs, reports, alerts, dashboards, and visualizations can be generated, as discussed below. For this reason, Splunk is used for searching the internet for Twitter and newspaper data by web scraping and creating the dashboards for analysing the online news media and Tweets. It is possible to expand the social media data collection to Facebook and YouTube data and other social media platforms using similar scraping techniques as Twitter. Web scraping eliminates the constraints imposed by the APIs, such as the limited number of records and restrictions that confine searches to recent date ranges only.

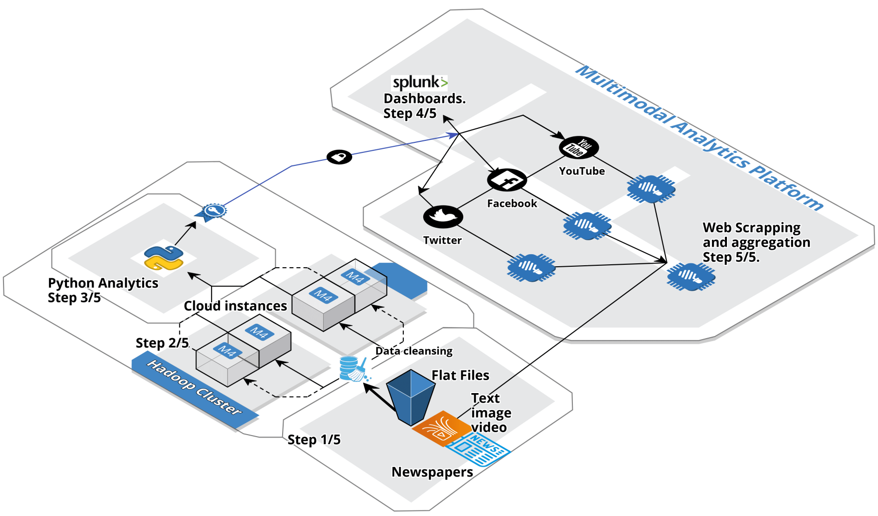


Figure 1: MAP deployment architecture in the cloud platform

The data ingestion process is as important as the analytics itself. In this case, users can search Twitter in real-time through an easy-to-use web interface in MAP. The newspaper data collection is an automated backend process. A Python process scheduled through the Linux Cron job incrementally updates the database with all the news articles published on that day from a list of newspapers. The newspaper data collection process is programmed once in 24-hour intervals (though the data collection times can be changed according to user needs). Each news article contains the following fields: date of publication, article title, author, publisher/name of the newspaper (in this case, *The Guardian*, *The Independent*, *The Evening Standard*, *The Metro* and *The Sun*), keywords, article summary, full text and URLs for articles, images, and videos. A Python process extracts keywords and summarises the news texts for each article. Users can search and create reports based on these two additional fields (i.e. keywords and the summary) which are not present in the original news articles. The newspaper scraping uses Python 3.7 Anaconda distribution along with Beautiful Soup 4.9.3 libraries (Nair, 2014) APIs to identify HTML elements (e.g. title, author, etc.) from each of the article URLs. The search page results for online news articles from 19 to 25 May 2020 is displayed in Figure 2a.

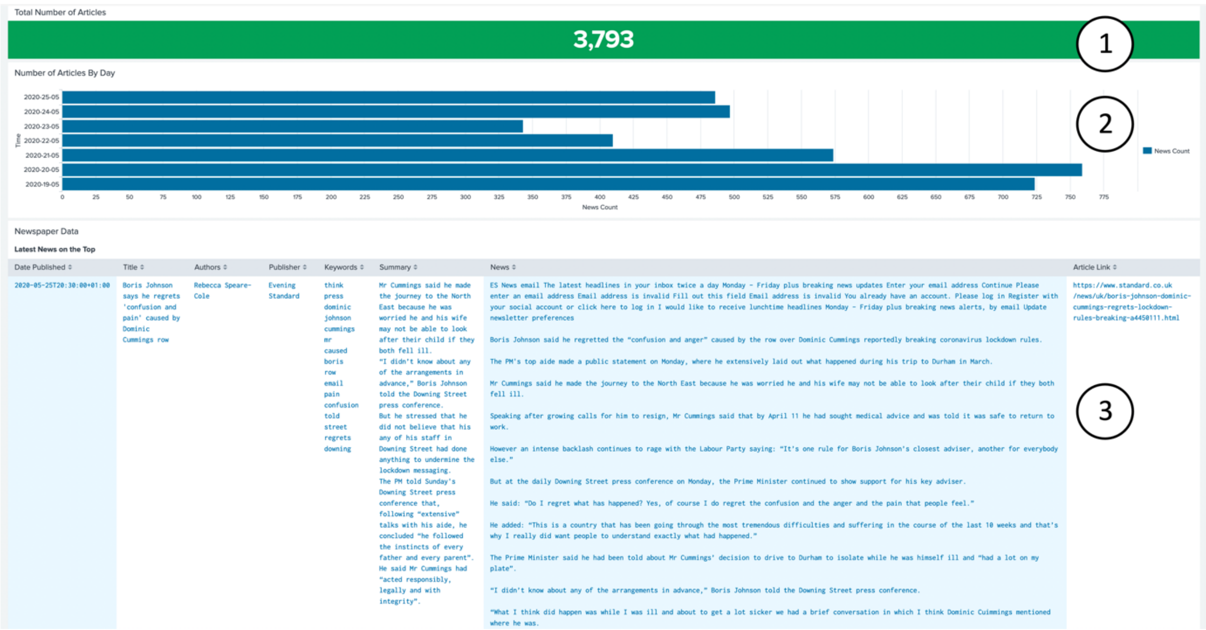


Figure 2a: Newspaper search page: number of articles [1]; counts of articles by date [2]; search results, i.e. date of publication, article title, author, publisher (newspaper name), keywords, summary, full text and URL [3]

Users need to undertake specific searches to generate Twitter data, given the logistics of securing this type of data. In the data discovery phase, users' keyword search looks for all possible matching Tweets in the Twitter database. Users can refine the search query using a date range and optional location filter. In the subsequent data collection phase, Python Beautiful Soup (Nair, 2014) APIs scrape the Twitter pages and save the data into the local file system as JSON formatted data. Each Tweet contains the meta fields along with actual Tweets that include the following metrics: Tweet text, hashtags, retweets, Tweet user name, Tweet, image, and video URLs, and the number of likes, replies, and retweets. Also, each record contains two other meta fields indicating if the Tweet is a reply to another Tweet or the Tweet is an original post. Each Tweet contains a primary key field as Tweet ID that uniquely refers to a Tweet post. For example, the results for the search "coronavirus" from 19 to 25 May 2020 is displayed in Figure 2b. The results show the total number of Tweets found in the search, the number of images, the number of Tweets for each day, the list of actual Tweets for the current search, and the list of Twitter searches which have been recently undertaken in MAP, as displayed in Figure 2b.

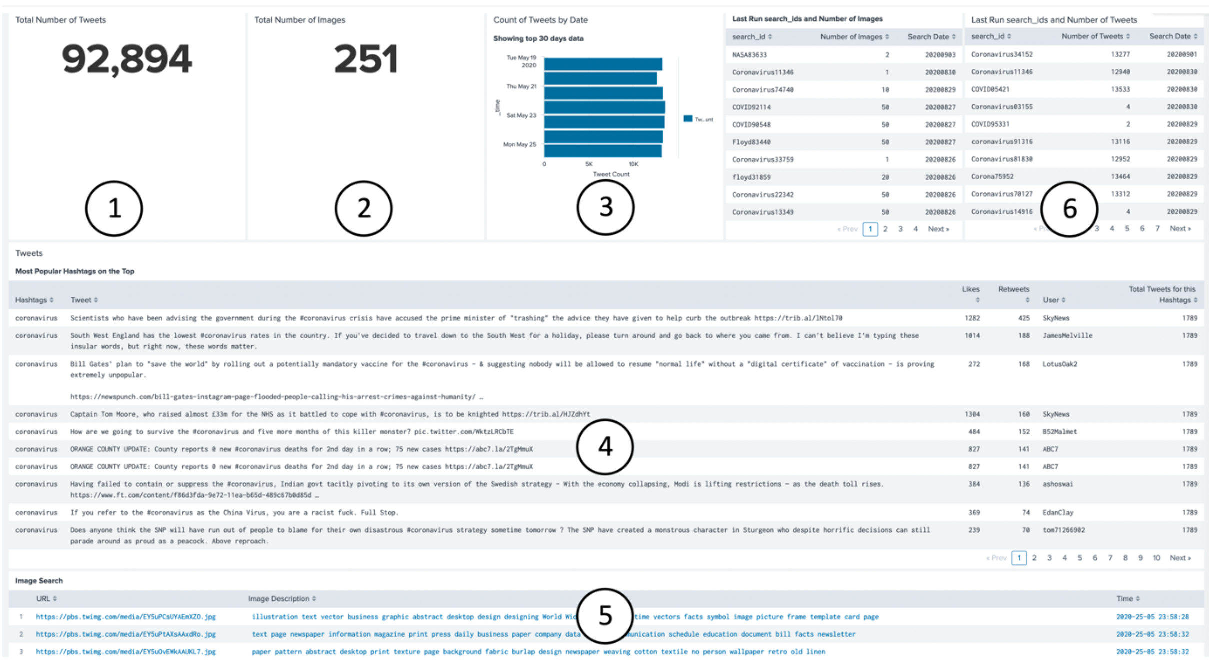


Figure 2b: Twitter search page: number of Tweets [1]; number of images [2]; counts of Tweets by date [3]; search results for Tweets: hashtag, Tweet text, likes, retweets, user, total number of Tweets for the hashtag [4]; Search results for Images: URL, image description, time [5]; latest searches for Tweets and images in MAP [6]

News and social media often rely on instantly capturing users' interest through various images and videos (e.g. Stöckl et al., 2020). MAP permits visual media to be searched, downloaded and thumbnails to be created for images and videos. However, MAP does not download and semantically annotate visual media data during the data discovery and collection phase described above. Downloading millions of images and video data for all the Tweets and news articles is economically unviable in the cloud infrastructure, given disk and internet bandwidth. Furthermore, meaningful reports are typically based on filtered datasets of relevant topics and date ranges. Hence, the user explicitly requests image and video searches for newspaper and Twitter data. Within MAP's media (i.e. images and videos) generation interface, users have complete control over the context and subject of the media that are retrieved. Users can refine media search criteria based on complex date rage, name of one or multiple newspapers, a count of media searched and analysed, and can apply a joint embedding of the texts, images, and videos using the menus provided in the interface. For example, the image search results page for the five online newspapers for 26 May 2020 are displayed in Figures 3a and 3b. The search page displays the image counts, the thumbnails, the search results in terms of the article title, the publisher, the date published, image description, the date data was extracted, the image link and the article link, as displayed in Figure 3a. In addition, the images can be viewed with the image link, as displayed in Figure 3b. The actual image can be displayed by clicking on the URL that opens the image in the original news article. The images can be downloaded for further processing.

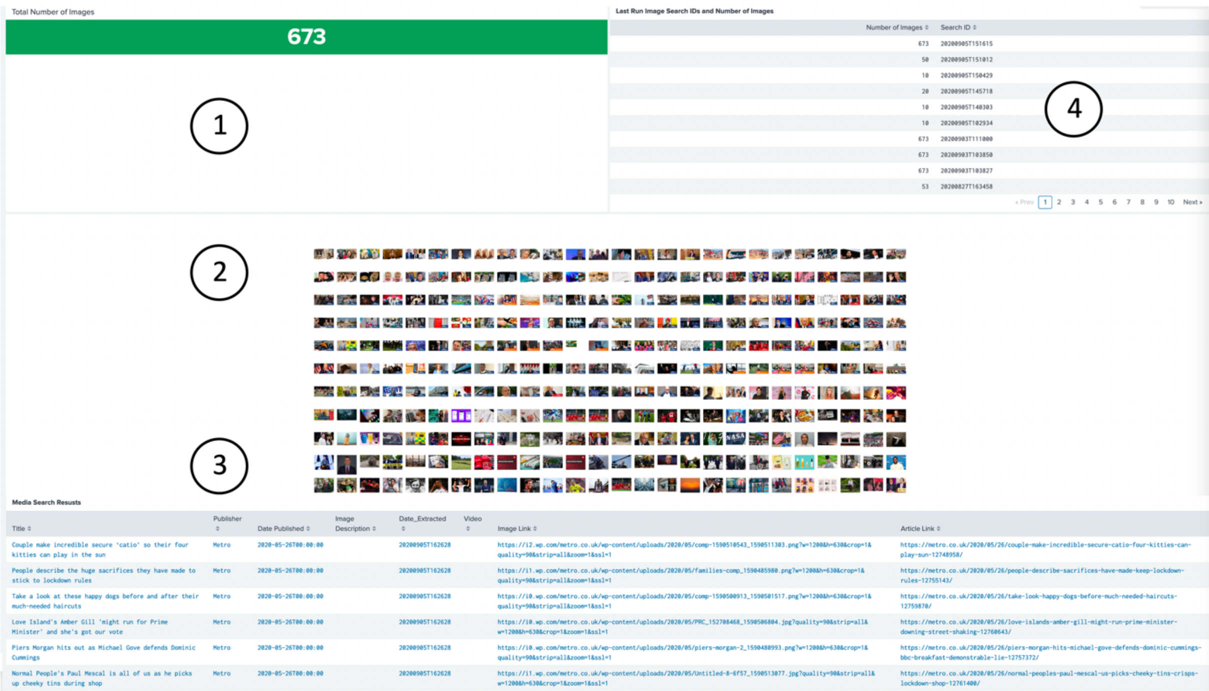


Figure 3a: Image search for online news articles: count [1]; thumbnails [2]; search results for title, publisher, date published, image description, date extracted, image link, article link [3]; recent searches with search IDs [4]

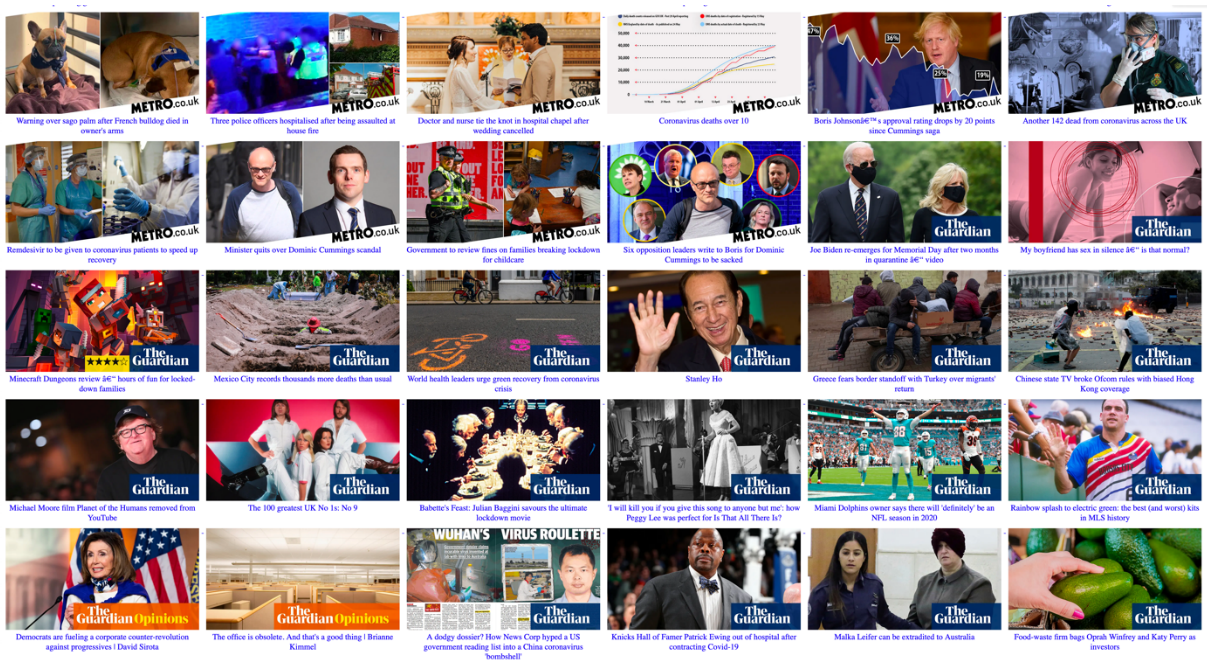


Figure 3b: Image search results: the images with image links

3.2 Indexing and Semantic Annotation

The case study addresses the multimodal correspondences between text and image data. In general, the method is based on studying the latent space of these media resources by producing annotations describing the media using NLP and image processing techniques which provide linguistic labels for visual content. MAP's interactive dashboard feature lets users select from possible combinations in the user interface to compare and build interactive reports. One of the key characteristics with MAP is that it computes NLP and machine learning tasks at low latency between search and reporting, providing a user-friendly interactive dashboard experience where search results are displayed with minimal delay. Each record processed in MAP is time-series data, i.e. data attached to a timestamp. The timestamp field for news articles and Tweets is the date of publication and the date when a Tweet is posted. For indexing on the collected data, MAP breaks the events based on the timestamp and ingests the data into the Splunk database. The most recent and frequently accessed data is also stored into a cache for even faster response. Therefore, in MAP, searching through 50,000 news articles, downloading and creating thumbnails of 50 images can be completed within approximately ten seconds, for example.

The images and videos resulting from users' explicit searches are sent to Clarifai computer vision APIs[7]. Clarifai is an external service to MAP that uses Convolutional Neural Networks (CNN) (Kalchbrenner et al., 2014; Kim, 2014; Krizhevsky et al., 2012), which is a class of deep neural networks to analyse and identify media and annotate them. The automated tagging for each image or video generates metadata describing the media in a set of English language descriptors which classify the images into different categories; for example, concepts including objects, themes, mood, and demographics (e.g. age, gender, cultural appearances of faces). MAP annotates thousands of images and videos filtered on a time range, headline, or location into textual descriptions. In addition, video recognition algorithms are used to tag the videos, grouped according to a time unit (e.g. second, minute, or hour). The confidence score (Koo et al., 2001; Mandelbaum and Weinshall, 2017) ranges between 0 and 1, denoting the level of certainty attached to the description of the image or video. The higher the score, the greater confidence of accuracy in the result, with a score of 1 indicating absolute certainty. For example, the Clarifai results for the screenshot from the video "Michael Gove says he has 'on occasions' drive to test his eyesight"[8] published in *The Metro* online news on 26 May 2020 are displayed in Figure 4. Michael Gove (Minister for the Cabinet Office, UK) made this statement in defence of Dominic Cummings (Chief Advisor to UK Prime Minister Boris Johnson) who had travelled from London to County Durham and visited Barnard Castle during a COVID-19 lockdown in the UK. As displayed in Figure 4, each concept has a confidence level associated with the result, making it possible to select descriptors which meet minimum confidence level requirements.

Concept	Probability	Concept	Probability
people	0.987	business	0.911
portrait	0.975	festival	0.909
one	0.972	outdoors	0.905
adult	0.963	building	0.897
man	0.944	light	0.884
administration	0.936	politician	0.874
leader	0.917	capital	0.843

Figure 4: Clarifai concepts and confidence Levels: Michael Gove while defending Dominic Cummings (video frame, published in *The Metro*, 26 May 2020)

3.3 Search and Interactive Reports

Multimodal analysis necessarily requires embedding techniques where text, image, and videos can be combined together (Mithun et al., 2018). At this stage, the approach involves constructing linguistic descriptions so that the resultant data subsequently can be jointly analysed through NLP techniques like n-grams, parts of speech, lemmatisation, sentiment analysis, similarity, and classifications tools. While the limitations of this method are recognised (see discussion section), this approach is possible, given that the core functionalities of MAP include the aggregation of multiple Twitter searches and news articles and removal of unwanted noisy data. MAP overcomes the limitations of existing social media analysis tools which cannot aggregate temporal modalities between news media and social media. Therefore, MAP has an embedded space to combine the modalities between text, image, and video across social and news media, batch data, and real-time data with multiple time-series properties, although the limitations of the current multimodal embedding techniques are recognised. The functionalities of MAP for analysing multimodal discourse trends are demonstrated in the case study involving online news media and social media reports about COVID-19 and George Floyd. Following this, the benefits and limitations of the approach are discussed.

4. The Case Study - COVID-19 and George Floyd

For the most part of 2020, COVID-19 has dominated social media and news media coverage. In comparison, intense public interest in events like George Floyd's death fluctuate, depending upon the latest incidences of black discrimination around the world. The focus of this case study is to track multimodal discourse trends about COVID-19 immediately before, during, and after a defining moment in history, such as the death of George Floyd which sparked international outrage, given the long history of police brutality towards black people (e.g. Peeples, 2020). In this case, George Floyd's death was captured on video, leading to protests around the world, organised by the Black Lives Matter political and social movement which advocates non-violent civil disobedience against police brutality and racially motivated violence and discrimination against black people[9]. At this time, the COVID-19 pandemic had heightened social and economic divides, with black and minority groups being much more likely to die with the disease due to poverty, overcrowded housing, and lower-paid and/or key worker roles (e.g. Evans, 2020). The case study examines how these tensions were played out in online UK newspapers and social media.

4.1. Data

The case study focuses on topics relating to COVID-19 and the death of George Floyd and the subsequent Black Lives Matter protests. Contrary to traditional tools that collect one-time static data, MAP continually collects news articles (at an interval) through real-time searching on the web. The Twitter database is also searched in real-time for a date range which can be in the past or the most recent time. Traditional tools that use Twitter APIs to query the Twitter database can only return 500 results on each search as the APIs are restricted to a small number of records on each call to Twitter. The free and paid Twitter APIs provides only 1% and 10% sample of all Tweets respectively which influences the results which are obtained (e.g. Pfeffer et al., 2018). The alternative method to collect data is by web scraping Twitter pages without using the APIs. However, Twitter, like other popular websites, use advanced scrape-detection software to protect sites from continuously scrapping, since crawling through the pages slows down the websites (Stevanovic et al., 2013; Zhuang et al., 2012). These limitations are overcome in MAP using multiple proxies and integrating the results from several searches. This data collection technique reliably collects the maximum amount of data from Twitter without data loss. The results presented here are based on an initial dataset of approximately 2 million Tweets, 76,000 news articles and 10,000 associated images.

Combining many search results on different topics, date range, and eliminating noisy unrelated topics is one of the key functionalities of the proposed framework. In this way, MAP can be used to examine responses towards the COVID-19 and the death of George Floyd through news and social media data lens. The data which is searched and filtered for this case study is based on the following topics: diagnostic testing, contact tracing, restrictions in the movements, international travel ban, vaccine trials, cancelled public events, school closure, workplace closure, general public awareness campaign (stay at home order), etc. As the media conversations changes over time, the analytical results are used to interpret the sequence of events and discourse trends.

Despite the initial downplaying of the dangers of the emerging pandemic, COVID-19 has continued to dominate news spaces. The news media coverage has elicited a range responses by internet users which largely accord with key news topics (e.g. Gozzi et al., 2020; Garfin et al., 2020), although there are temporal deviations to these patterns. For example, Gozzi et al. (2020) show that "collective attention was mainly driven by media coverage rather than epidemic progression, rapidly became saturated, and decreased despite media coverage and COVID-19 incidence remaining high". However, as people read more news because of COVID-19, social media trends followed the news reporting, resulting in a 'social media infodemic' (Cinelli et al., 2020). Following the first reported death by China, WHO's first situation report and the subsequent declaration of global emergency, made COVID-19 the dominant topic in social media from the first week of February 2020. The virus has spread to all parts of the world and the social divides had become evident during the period 19 May 2020 to 2 June 2020 when this case study takes place. However, as we will see, public attention also turned to issues of racial discrimination and violence following the death of George Floyd.

4.2 Results

MAP generates dashboards to visualize large datasets with an ability to customize, filter, or combine multiple datasets to create reports using Splunk and Python. This section provides the results of the text and image analysis for the five UK online newspapers and Twitter for key word searches for 'coronavirus', 'Floyd' and "Black Lives' for these time periods:

1. 19-25 May 2020 - One week before George Floyd's death, henceforth 'Week Before'
2. 26 May 2020 (British Summer Time) - George Floyd's death, henceforth 'The Day'
3. 27 May 2020 - 2 June 2020 - One week after George Floyd's death, henceforth 'Week After'

The news media responses for the three time periods, Week Before, The Day, and Week After, are based on more than 10,000 news articles published in the five UK newspapers during this time period. In addition, Twitter searches for coronavirus, Floyd, and black lives resulted in 208,500 Tweets for these keywords. The distribution of newspaper articles and Tweets for these search terms for the three time periods are displayed in Table 1

Time Period	**Newspaper Articles**			**Number of Tweets**		
	Coronavirus	George Floyd	Black Lives	Coronavirus	George Floyd	Black Lives
Week Before 19-25 May 2020	1,900	0	0	79,575	0	10,063
The Day 26 May 2020	297	2	1	13,665	13,666	6,672
Week After 27 May-2 Jun 2020	1,514	590	42	82,013	72,366	76,135

Table 1: Newspaper articles and Tweets counts for 'coronavirus', 'Floyd' and 'Black Lives' for Week Before, The Day, and Week After

The results for classifications and sentiment analysis, with examples of the cluster and tag cloud analysis for online news and Twitter around the topics of COVID-19, George Floyd and Black Lives are discussed below. These first set of results are based on language analysis. Following this, the image analysis and multimodal integration between language and visual features are investigated, in order to critically examine the methodology involving linguistic descriptors for visual images. In each case, the visualisations are produced using MAP, based on the search criteria which have entered: that is, time period, data source (newspapers or Twitter) and keywords.

A. Classifications

The volume of newspaper articles about coronavirus in the week proceeding George Floyd's death (1,900 articles) decreased during the week following his death (1,514 articles) as attention diverted to this pivotal event (see Table 1). Nonetheless, the classification results (see Figure 5) showed that the news articles about coronavirus were primarily concerned with politics and wellness in both weeks. For example, politics and wellness accounting for approximately 30% and 10% of articles respectively in the Week Before, and this trend continued in the Week after. However, the number of articles about coronavirus which were classified as ‘black voices’ increased slightly (from 4 articles to 13 articles), indicating growing interest between coronavirus and the Black Lives Matter movement. We may also see from the 3D cluster diagram (Figure 6), that the news articles classified as ‘politics’ and ‘wellness’ are more widely dispersed (and hence disparate) compared to articles that have been classified as ‘entertainment’ which form a distinct cluster.

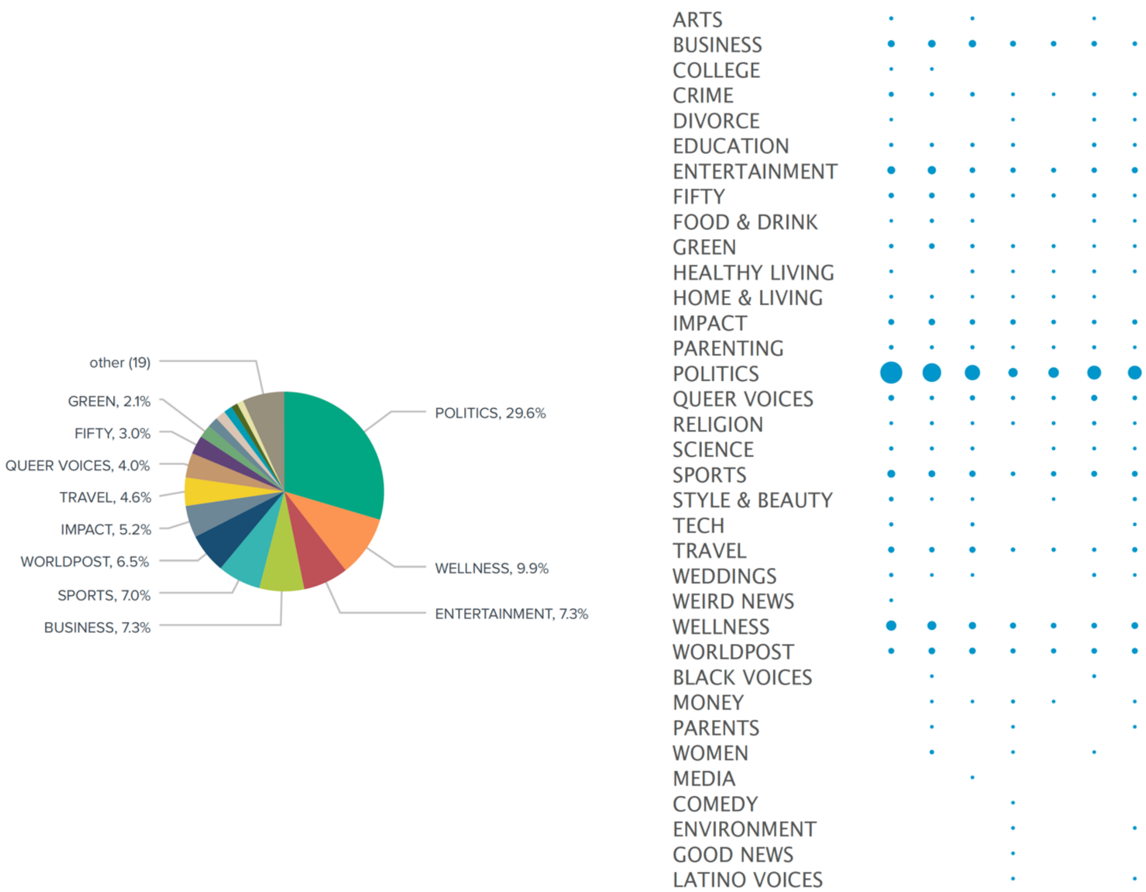


Figure 5: Classifications of Newspaper articles with the keyword ‘coronavirus’ for Week Before. In the punchcard chart (right), each column represents one day of the Week Before and the circles represent the relative volume of articles.

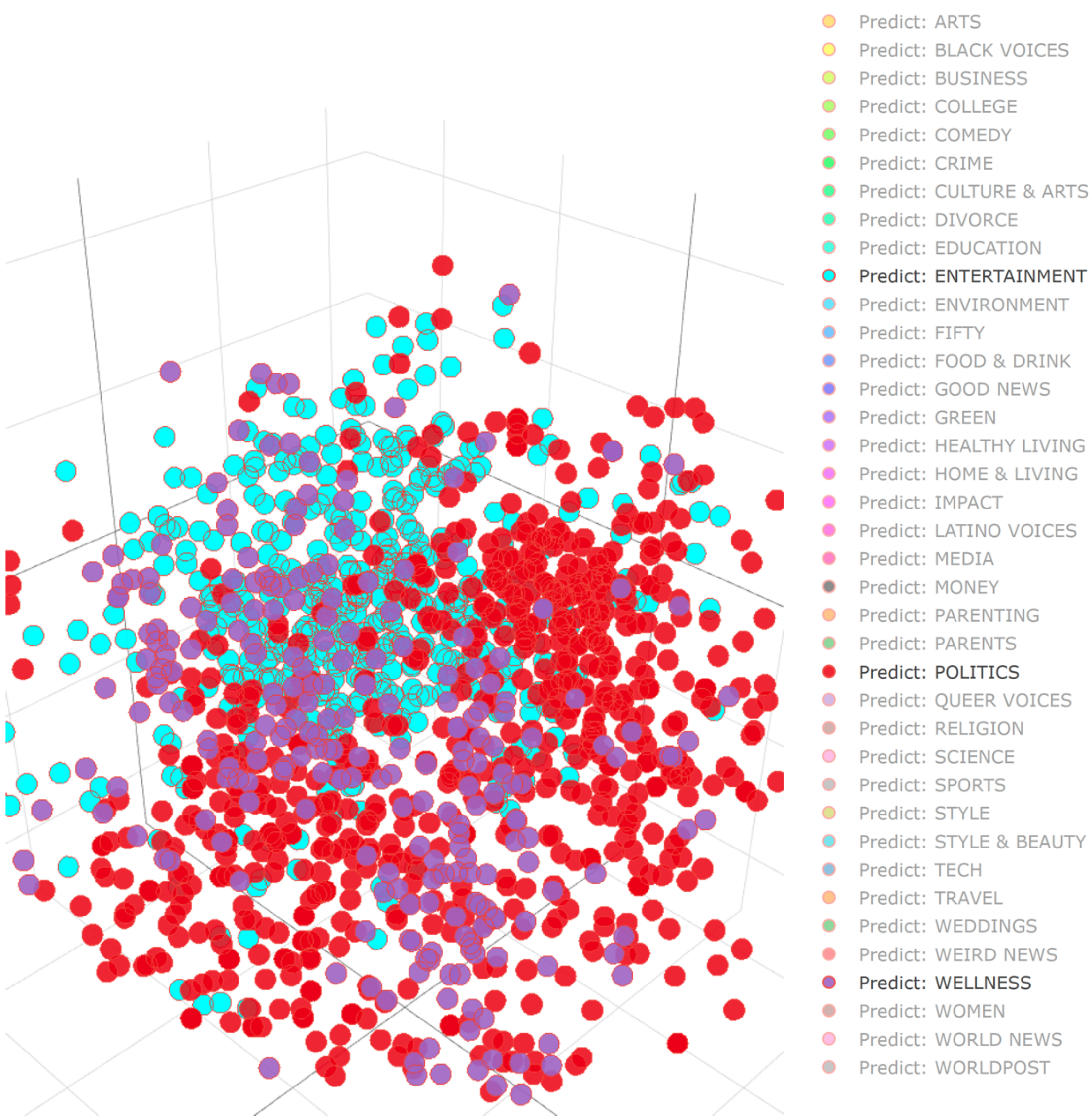


Figure 6: Clusters for news articles classifications 'politics', 'wellness' and 'entertainment' in the Week Before. The 3D scatterplot visualization uses Principal Component Analysis (PCA) to reduce high-dimensional feature vector into 3 axes (x, y and z).

The results for coronavirus articles may be contrasted with the articles about George Floyd. That is, in the week following George Floyd's death, the news articles with the keyword "Floyd" (590 articles) were primarily concerned with politics (43.1%), black voices (15.5%) and crime (10.7%), accounting for 70% of all the articles about Floyd. In addition, 42 articles focussed on "Black Lives " with classifications politics (45.2%) and black voices (31.0%). In summary, approximately 40% of newspaper articles in the five newspapers focused on coronavirus, George Floyd and Black Lives during this time, with an increasing focus on politics, black voices and crime. The word cloud for newspaper articles for Floyd on The Day is displayed in Figure 7a. The reporting is factual, and is largely concerned about the circumstances of George Floyd's death. However, the word cloud for newspaper articles about George Floyd in the Week After in Figure 7b reveals the shift to reporting about reactions to the event in relation to protests, riots, violence, and the arrest of the police officer.

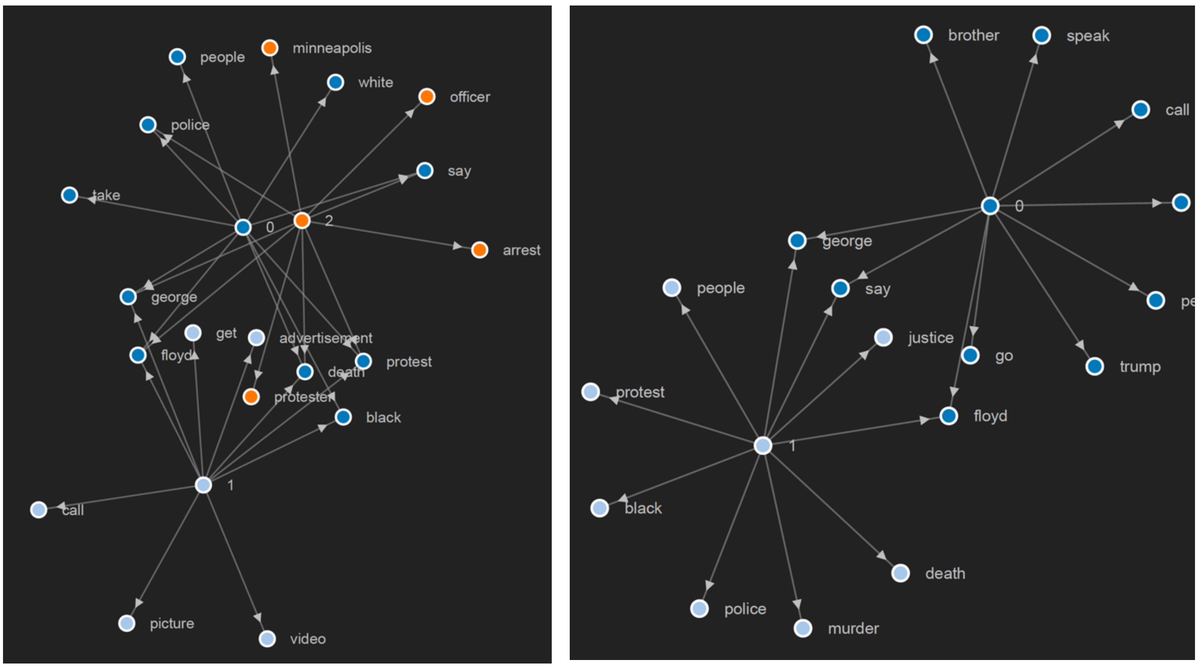
(a) Newspaper paper articles on the Day **(b) Newspaper articles in the Week After**

Figure 7: Word clouds for Newspaper Articles on George Floyd

The reactions in Twitter to the coronavirus and Floyd showed differences to mainstream news reporting. For example, tweets about coronavirus before George Floyd's death (79,575 tweets) were primarily concerned with healthy living (28.7%), politics (20.7%), and wellness (14.9%), showing that health was the major concern of the public (as opposed to politics in online news). On the day that George Floyd died, tweets about coronavirus (13,665 tweets) showed a similar pattern, with a focus on health. Following George Floyd's death, tweets about coronavirus did not abate (82,013 tweets) (unlike news reporting) and the concerns with healthy living (23.7%) , politics (14.3%), and wellness (9.6%) remained, but with increased interest in black voices (3.0%). In other words, the public continued to focus on health-related issues in Twitter, whilst turning attention to Black Lives Matter.

On the day that George Floyd died, there were 13,666 tweets with the keyword "Floyd". The classification results showed that the tweets were primarily about politics (33.3%), black voices (26.5%) and crime (15.0%). In the week after, the tweets about George Floyd (72,366 tweets) moved firmly in the realm of politics (60.4%), black voices (18.7%) and crime (4.6%) The volume of tweets about George Floyd was less than the number for coronavirus (72,366 tweets compared 82,013 tweets respectively) but the discrepancy was far less than the difference in newspaper articles about these two topics. In other words, the public reaction to the death of George Floyd nearly matched the level of concern about coronavirus on social media in the Week After. Subsequent to the defining moment of George Floyds' death, protest marches for Black Lives Matter were organised around the world. The Black Lives Matter movement gained momentum amidst the outrage caused by the killing of George Floyd, supported through social media.

Further differences between newspaper reports and public reaction on Twitter are evident in K-means clustering results, displayed in Figure 8. Clustering uses a Topic Modelling algorithm that identifies and groups homogeneous terms based on the topic similarities in semantics. Clustering results show that the newspapers articles concerned with Floyd during the Week After show interacting clusters of terms, with the common terms "George", "Floyd", "death" and "protest" in three clusters being linked to the terms "black" and "police" in two clusters, as displayed in Figure 8a. On the other hand, the Twitter clusters are distinct, with terms "George", "Floyd" and "say" interacting with the terms "brother", "call", "go", "peace", "speak", "trump" and "violence" in one cluster and "black", "death", "justice", "murder", "people", "police" and "protest" in another cluster. In this way, we can see that members of the public specifically link wider political issues involving US President Donald Trump and murder to the killing of George Floyd in Twitter.



(a) Newspapers clusters for Floyd **(b) Twitter clusters for Floyd**

Figure 8: Cluster Analysis for newspapers and Twitter in the week after George Floyd's death

B. Sentiment Analysis

The overall sentiment in newspaper articles about coronavirus tends towards neutral and low negative values, as positive and negative sentiment provide counter balances to each other, given the different aspects being reported (e.g. death counts, government initiatives, and good feel news stories). Also, at the beginning of the outbreak, due to a lack of scientific information, public opinion was greatly divided about the long term impact of COVID-19. Therefore, a large variance in sentiment polarity with a higher level of contradiction resulted in the overall sentiment values being close to neutral. A similar trend is found in Twitter posts about coronavirus where members of the public debate different aspects of the pandemic with divided opinions. However, the same cannot be said about the online reporting and tweets about George Floyd. As displayed in Figure 9a, the sentiment score for online newspaper articles about Floyd is -0.513, and this increased to -0.643 in articles during the week following his death. The sentiment coordinates in Figure 9b show the distribution of the weightings across the different articles, with the average sentiment values exceeding -0.5 in terms of value.

Figure 9a: Sentiment value for newspaper articles about George Floyd on 26 May 2020. The aggregated average sentiment of the articles remained negative. The pie chart shows the distribution of sentiment in three categories as positive, negative and neutral.

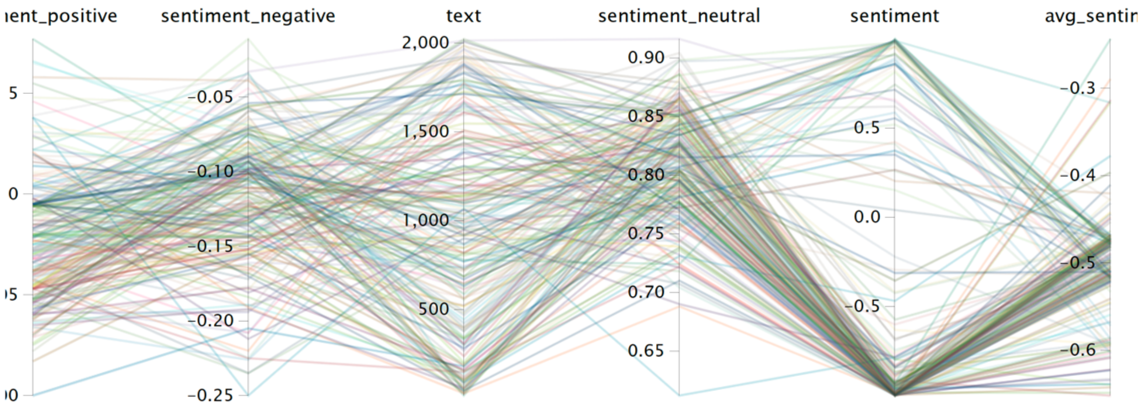


Figure 9b: Sentiment coordinates (positive, negative, neutral, and overall score) for newspaper articles about George Floyd on 26 May 2020

The sentiment analysis of Tweets about George Floyd has a negative score of -0.381 on the day and -0.158 during the week which follows. This somewhat surprising results perhaps reflect the nature of the postings on Twitter which include calls to rally, rather than the actual sentiment expressed in the tweets (e.g. "Mayor Jacob Frey: Justice for George Floyd - Sign the Petition! http://chng.it/R7ccNpSQ via @Change"). The sentiment for Black Lives has similar values, with low levels of negative sentiment in the Week Before and Week after (e.g. -0.091 and -0.060 respectively). However, the outcry of negative sentiment rose significantly on The Day when George Floyd died (-0.268), which galvanised members of the public to take action which resulted in protests around the world.

C. Embedding Language Modality with Visual Modality

MAP introduces a new model for multimodal fusion between language (text), visuals (image), and textual metadata. MAP supports joint intermodal representation across a large number of images and associated texts. The multimodal embedding process is explained using an article published in *The Independent* newspaper on 26 May 2020 following George Floyd's death. The article is titled: "*As a black man, who's watched white Republicans fake outrage over Biden saying 'You ain't black', I need you to know this"*[10]. The image (Figure 10) originally associated with the article (which has since been replaced with a video) is converted to textual descriptions by the Clarifai image classifiers, with the following results: *administration, people, business leader, politician, man, chair, home, football, portrait, meeting, democracy, league, leadership, banking, intelligence, flag, parliament, military achievement*.

Figure 10: Image associated with a news article converted to textual descriptions

Classification Labels
arts & culture, black voices, business, college, comedy, crime, education, entertainment, fifty, good news, green, healthy living, impact, Latino voices, media, parents, politics, queer voices, religion, science, sports, style, taste, tech, the world post, travel, weird news, women, world news, world post, parenting, home & living, environment, weddings, divorce, food and drink, money, wellness

Table 2: Classification labels

MAP classifies the news text and associated images into one of the 38 classification labels. The full list of classification labels is displayed in Table 2. Text and image classification results are combined in accordance with the confidence scores and individual weights applied to the image and text. In addition, the following metadata are extracted from the text: subjectivity, sentiment polarity, and word count. The subjectivity score ranges between 0 (completely objective) to 1 (completely subjective). The sentiment polarity score is computed between -1 (completely negative) to +1 (completely positive). Word count is the number of total words in each document (Tweet or news article). Metadata are converted to a numerical feature vector and integrated into the text embedding (see Figure 11). The outcome of the joint modalities from text and metadata is combined with image modality after applying weights on each modality. Users can select the weight scores for text and images (see Figure 12). The confidence score is the probability value of an image or text being classified into a particular label. For example, text in the article is classified as 'black voices' with a probability of 30%, and the associated image descriptions are classified as 'politics' with a probability of 70%. As equal weight is applied for text and image (50% each), the combined classification label is computed as ‘politics’ (see Figure 13).

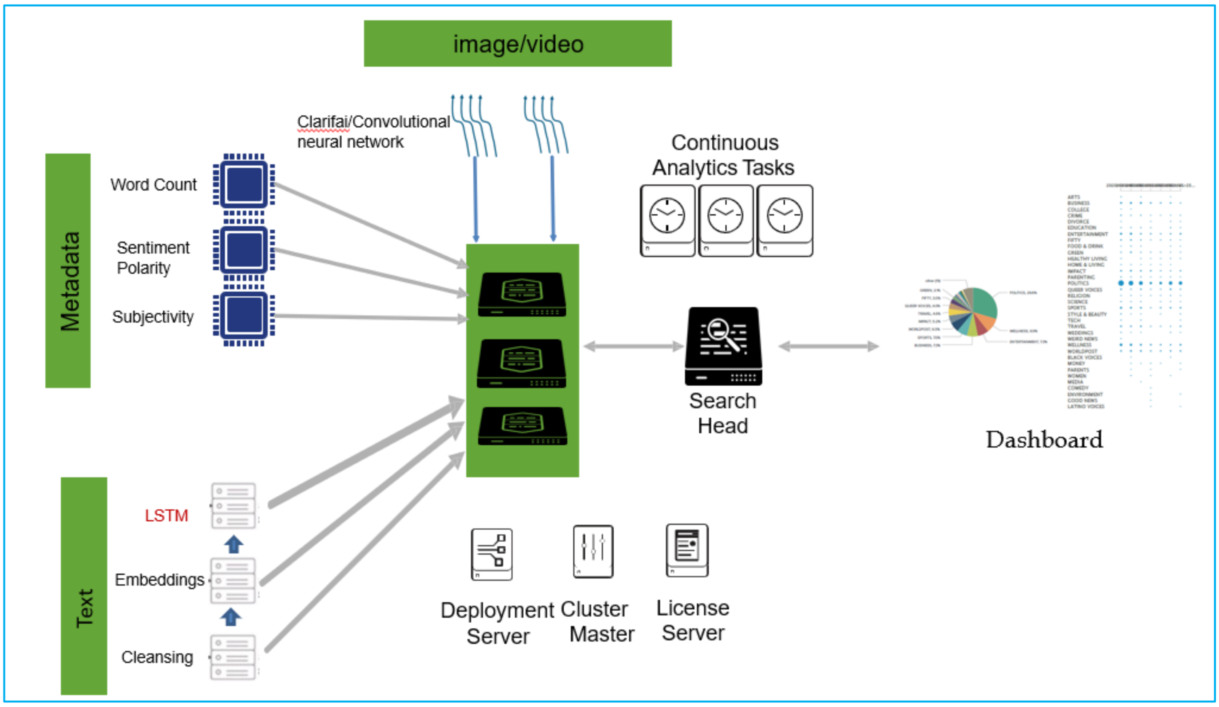


Figure 11: Language features are decoded using Long Short-Term Memory (LSTM), a class of Recurrent Neural Network, Visual features are encoded with Convolutional Neural Network (CNN). Numeral features such as sentiment polarity, subjectivity, and word count are extracted from the text. A simple multilayer perceptron integrates modalities. Users make searches on the data with combined modality to create reports.

Figure 12: User Interface for multimodal classification. The user is able to select a weight for image and textual features

Date	feature_img	image	title	Summary	Text_Predict	Image_Predict	Predict
2020-05-26	https://static.independent.co.uk/s3fs-public/thumbnails/image/2020/05/25/19/trump-biden-mask.jpg	administration people business leader politician man chair home football portrait meeting democracy league leadership banking intelligence flag parliament military achievement	As a black man who's watched white Republicans fake outrage over Biden saying 'You ain't black', I need you to know this	Nonetheless, the way Biden worded the comments hurt some black voters, who have carried the Democratic Party for decades and handed Biden his primary wins this year. Many of the reactions to Biden's comments were measured, advocating that the former VP should do more to assure the black community he isn't taking them for granted. Some used this as an opportunity to increase pressure on Biden to pick a black woman as his running mate. If you're black and reading this, this is why your vote should not be deterred by Biden's comments. President Trump smeared the first black president as being born in Kenya, spent his presidency trying to undo his legacy, and is now leading calls to imprison him.	BLACK VOICES	POLITICS	POLITICS

Figure 13: Text and image classification labels are *Black Voices* and *Politics*. A combined classification label is computed between image and text modalities as *politics*

The integration of the language and image results impact on the overall classifications of the newspaper articles and Twitter data. For example, we can compare the results for the classification of newspaper articles with the keyword ‘coronavirus’ for the week before George Floyd died for (a) text only (see Figure 5) and (b) text and image results displayed (see Figure 14). The classification results differ, with a decrease in the articles classified as politics (from 29.6% to 19.0%) and an increase in articles classified as travel (from 4.6% to 12.5%) , sports (7.0% to 8.7%) and entertainment (7.3% to 7.7%). The results demonstrate that the inclusion of image data (as textual descriptors) significantly influences the classification outcome, suggesting that the integration of text and image modalities has the potential to increase understanding of the meanings made in online communications. The larger implications of these findings are evident – namely that it is imperative to develop robust methodologies for image analysis in order to understand discourse patterns across media platforms. We return to this point in the concluding sections of the paper.

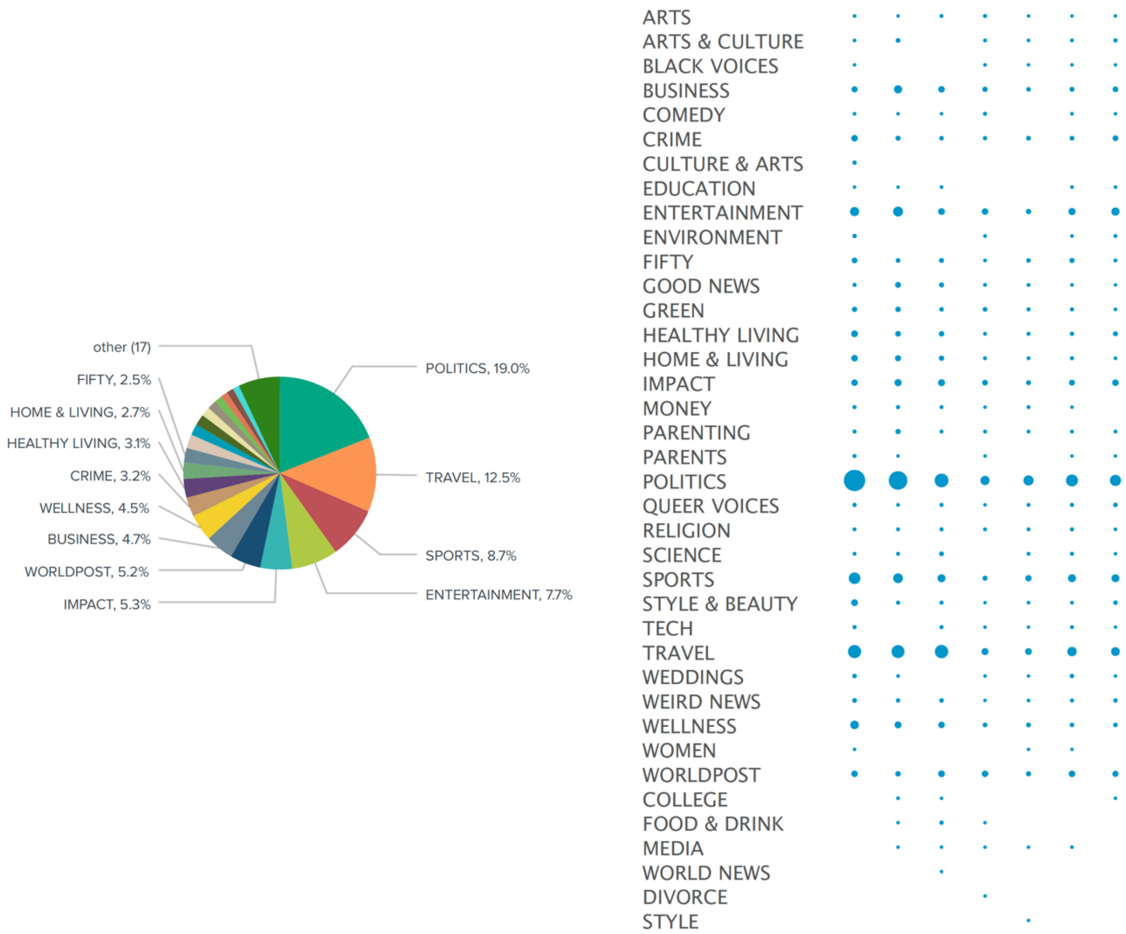


Figure 14: Classifications of Newspaper articles (text and image) with keyword 'coronavirus' for Week Before

D. Identifying the key topics

Information extraction is a method which accepts language as input and provides prearranged sets of elements such as places, people, organization after entity disambiguation. Compared to news articles which are professionally written and diligently edited, Tweet posts are short and in some cases ambiguous without knowledge of the context. In language analysis, making sense of Tweets involves a processing pipeline of language recognition, tokenization, Part of speech and Named Entity tagging to improve informativeness by extracting domain knowledge in terms of *who, what, when, where, and how*, in order to understand key information. Conventional analytics systems breaks the records into words and place them together either by putting all entities into a single bucket (one-hot vector) or separate buckets for each record (bag of words) (Sethy and

Ramabhadran, 2008; Wang et al., 2018). A collection of words without relation to other words do not retain their semantic meaning. Therefore, MAP adopts a different approach by identifying up to 19 Named Entity Recognition (NER) and 37 Parts of Speech (POS) tags, while maintaining explicit relationships between the tags. As shown in the ribbon chart (Figure 15), NER tags and the correlation within them, extracted from Tweets, before and after George Floyd incidence, noticeably captures the shift in focus as the conversation tilted towards the USA, Black Lives and president Trump from COVID issues in China, Brazil, USA and the UK. However, the analysis of the image tags which aggregates the generic descriptions of the thousands of images failed to reveal such shifts in focus (see Figure 16), again pointing to the need for more sophisticated image processing techniques.

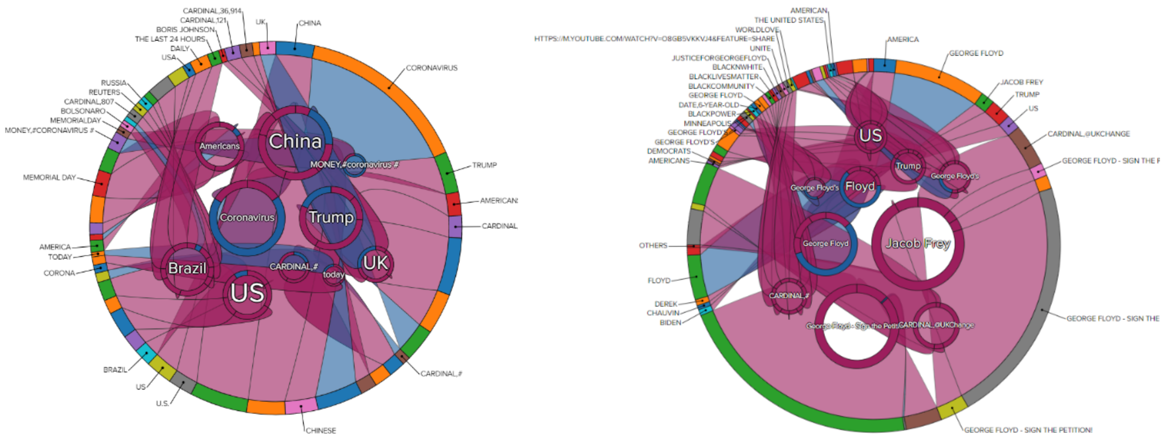


Figure 15: Named Entity Recognition for Twitter texts before and after George Floyd's death

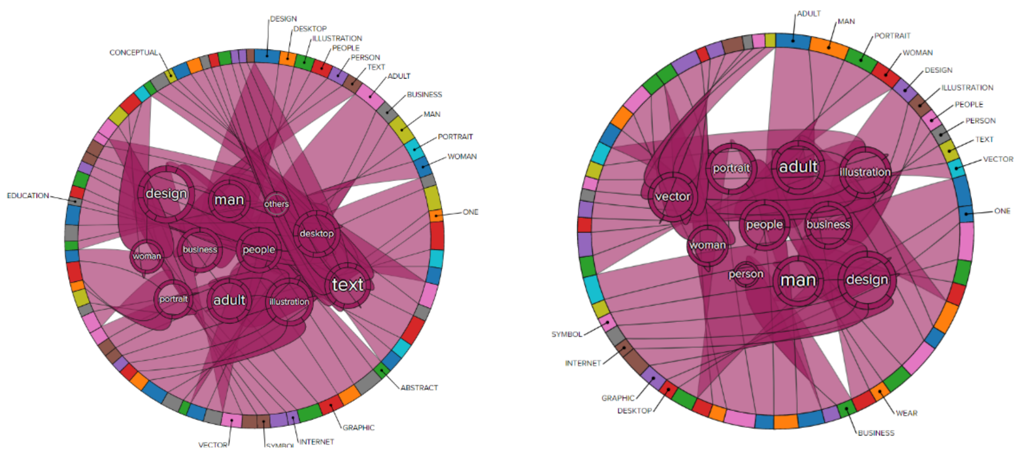


Figure 16: Named Entity Recognition for Twitter images before and after George Floyd's death

5. Discussion

As evidenced by this study, computational approaches to the multimodal analysis of large datasets is a complex undertaking involving multiple processes, including data collection, data storage, data cleansing, data analysis of language and image, multimodal integration of the results, and subsequent classifications and visualisation of the results to understand patterns and trends. In this case study, we collected online newspaper articles and Twitter data and applied Natural Language Processing tools and image processing tools and integrated the results in order to interpret differing reactions to COVID-19, George Floyd and Black Lives Matter. In doing so, the case study revealed that despite the stress, anxiety, and damage caused by the COVID-19, the debate around George Floyd's death and racism and prejudice resulted in the highest collective level of negative reactions in online news media and social media, compared to the polarising and highly differentiated response to the pandemic, at least in the period under consideration (i.e. 19 May to 2 June 2020). In addition, the results revealed the primary concerns about COVID-19 related to health and well-being in Twitter, compared to politics in the online newspapers.

However, the key issue arising from this case study involves the development of a theoretically-informed methodology for visual analysis which can be integrated with the linguistic analysis. Current image recognition algorithms depict images in terms of abstract captions, which involves a significant reduction in meaning which semiotic analysis is able to capture. Moreover, in our case study, the Clarifai image captions do not identify the celebrity names (although these models exist) and the place of the event. For example, the image tags for Donald Trump are: *administration, people, business, leader, politician, man, democracy leadership, parliament.* These tags are then categorized into a single classification label as '*politics'.* This result is useful for certain purposes, as demonstrated by the classification and cluster analyses presented in this study. However, further investigation is needed to expand our approach to incorporate more sophisticated image processing algorithms to identify personalities, objects, places, and events more precisely, in order to establish relations between key participants, processes, and circumstances across linguistic and visual analyses. Indeed, digital humanities researchers have started to explore how image processing can be combined with interpretative frameworks (e.g. Arnold and Tilton, 2019; Munster and Terras, 2020; Wevers and Smits, 2020). Further to this, the current scope of image recognition needs to be extended beyond classification to analyse sentiment, topic modelling, and other dimensions which are captured through the linguistic analysis. Despite these significant issues, this study has involved the development of an online platform with facilities for integrating tools and methodologies for multimodal approaches to big data analytics to reveal what can be achieved computationally at this point in time. The challenge remains to develop new tools and approaches to address the evident shortcomings of what can be achieved at present.

6. Conclusion

This work introduced a novel Multimodal Analysis Platform (MAP) that is capable of deriving critical attributes from multiple modalities (e.g. language, visual images, and metadata) with the aim of capturing patterns and trends within and across news and social media outlets. The platform provides facilities for multimodal integration of disparate datasets across social and news media, as demonstrated through the analysis of articles in five UK newspapers and Twitter data about COVID-19, George Floyd's death, and the ensuing Black Lives Matter protests. The multimodal methods for news and Tweet classifications use topic modelling algorithms to provide clusters of similar terms that are syntactically identical or most frequently co-occurring within the same record. In addition, entity recognition and tagging were introduced to eliminated noisiness in social media data. Multimodal classification, entity disambiguation and clustering helped formally characterize the cohesion, discord, or reciprocity within different sets of arguments about COVID-10 and George Floyd. For example, sentiment analysis reveals higher than usual levels of contradiction in general public opinion about the COVID-19 pandemic compared to the negative reaction to George Floyd's death. Empirical observations reveal how multimodal analysis of news reports and social media data can reconstruct the impact of the pandemic and acts of violence on society. However, these same empirical techniques show that the need to develop more sophisticated image processing algorithms if these multimodal trends are to be really understood.

From this perspective, the study reveals the necessity of using a multimodal theoretical framework to integrate the various computational tools into a robust methodological approach for investigating information in online media. The integration of multimodal framework with computational techniques will be a revolutionary technological innovation because the analysis will be informed by foundational principles about how human communication is organised to fulfil certain functions according to context. As Halliday (2009) has shown, human communication can be modelled as 'system' and 'text' – that is, as a set of systems from which options are selected according to the requirements of the situation. However, current computational techniques lack the foundational principles for formulating these systems and drawing out the resulting patterns on a consistent basis. This is precisely the knowledge which multimodal frameworks can harness. This meeting of theory and technology will remain a critical research agenda for the foreseeable future.

Funding

This research was funded by the Center for Interdisciplinary Research (ZIF), Bielefeld University, Germany 2019/2020.

References

Arnold, T., Tilton, L., 2019. Distant viewing: analyzing large visual corpora. Digital Scholarship in the Humanities 34 (Supplement 1), i3-i16.

Atrey, P. K., Hossain, M. A., El Saddik, A., Kankanhalli, M. S., 2010. Multimodal fusion for multimedia analysis: a survey. Multimedia Systems, 16(6), 345–379.

Bateman, J. A., 2017. Triangulating transmediality: a multimodal semiotic framework relating media, modes and genres. Discourse, Context & Media, 20, 160-174.

Bateman, J. A., 2019. information design and multimodality: new possibilities for engagement across theory and practice. Information Design Journal, 25(3), 249-257. doi:https://doi.org/10.1075/idj.25.3.02bat

Bateman, J., Wildfeuer, J., Hiippala, T., 2017. Multimodality Foundations, Research and Analysis – A Problem-Oriented Introduction. Berlin: De Gruyter Mouton.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., et al., 2020. The COVID-19 Social Media Infodemic Scientific Reports, 10 (1). doi:https://doi.org/10.1038/s41598-020-73510-5

Dumas, B., Lalanne, D., Oviatt, S., 2009. Multimodal interfaces: a survey of principles, models and frameworks, in: Lalanne, D, Kohlas, J. (Eds.), Human Machine Interaction. Springer, New York, pp. 3–26.

Evans, M. K., 2020. Covid's color line — infectious disease, inequity, and racial justice. The New England Journal of Medicine, 383, 408-410. DOI: 10.1056/NEJMp2019445

Garfin, D. R., Silver, R. C., Holman, E. A., 2020. The novel coronavirus (COVID-2019) outbreak: amplification of public health consequences by media exposure. Health Psychology, 39(5), 355-357. doi:http://dx.doi.org/10.1037/hea0000875

Gozzi, N., Tizzani, M., Starnini, M., Ciulla, F., Paolotti, D., Panisson, A., Perra, N., 2020. Collective response to media coverage of the COVID-19 pandemic on Reddit and Wikipedia: mixed-methods analysis. Journal of Medical Internet Research, 22(10). doi:10.2196/21597

Guinness, D., Cutrell, E., Morris, M. R., 2018. Caption crawler: enabling reusable alternative text descriptions using reverse image search, in CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, New York, pp. 1-11. https://dl.acm.org/doi/10.1145/3173574.3174092

Halliday, M. A. K., 2009. Collected Works of M.A.K. Halliday (10 Volumes). London and New York, Continuum.

Hoel, E. G., Heng, W.-L., Honeycutt, D., 2005. High performance multimodal networks, in: Medeiros, C. B., Egenhofer, M. J. and Bertino, E. (Eds.), Advances in Spatial and Temporal Databases. SSTD 2005. Lecture Notes in Computer Science, Vol. 3633. Springer, Berlin, pp. 308–327.

Ideya Ltd. (2018). Social Media Monitoring Tools and Services Report Public Excerpts 2018. http://ideya.eu.com/images/SMMTools%20Excerpts%202018%20Final.pdf

Kalchbrenner, N., Grefenstette, E., Blunsom, P., 2014. a convolutional neural network for modelling sentences, in: Toutanova, K., Wu, H. (Eds.), Proceedings of the 52nd Annual Meeting of the

Association for Computational Linguistics: Volume 1: Long Papers. Association for Computational Linguistics, Stroudsburg, Pennsylvania, pp. 655-665.

Kim, Y., 2014. Convolutional neural networks for sentence classification, in: Moschitti, A., Pang, B., Daelemans, W. (Eds.), Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Stroudsburg, Pennsylvania, pp. 1746-1751.

Koo, M.-W., Lee, C.-H., Juang, B.-H. (2001). Speech recognition and utterance verification based on a generalized confidence score. IEEE Transactions on Speech and Audio Processing, 9(8), 821–832.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Pereira, F., Burges, C. J. C., Bottou,L., Weinberger, K. Q (Eds.), Advances in Neural Information Processing Systems *25*. Curran Associates Inc, Red Hook New York, pp. 1097–1105.

Larios, E., Zhang, Y., Yan, K., Di, Z., LeDévédec, S., Groffen, F., Verbeek, F. J., 2012. Automation in cytomics: a modern RDBMS based platform for image analysis and management in high-throughput screening experiments, in: He, J., Liu, X., Krupinski, E. A., Xu, G. (Eds.), International Conference on Health Information Science. Springer, New York, pp. 76–87.

Mandelbaum, A., Weinshall, D., 2017. Distance-based confidence score for neural network classifiers. arXiv preprint arXiv:1709.09844.

Maynard, D., Dupplaw, D., Hare, J., 2013. Multimodal sentiment analysis of social media. doi:https://eprints.soton.ac.uk/360546/1/arcomem.pdf

Mithun, N. C., Li, J., Metze, F., Roy-Chowdhury, A. K., 2018. Learning oint embedding with multimodal cues for cross-modal video-text retrieval. Paper presented at the ICMR '18: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, Yokohama, Japan. Association for Computing Machinery, New York, pp. 19-27.

Munster, S., Terras, M., 2020. The visual side of digital humanities: a survey on topics, researchers, and epistemic cultures. Digital Scholarship in the Humanities, 35(2), 366-389.

Nair, V. G., 2014. Getting Started with Beautiful Soup. Birmingham UK: Packt Publishing Ltd.

O'Halloran, K. L., Tan, S., Pham, D.-S., Bateman, J., Vande Moere, A., 2018. A digital mixed methods research design: integrating multimodal analysis with data mining and information visualization for big data analytics. Journal of MIxed Methods Research, 12(1), 11-30. doi:10.1177/1558689816651015

O'Halloran, K. L., 2020. Multimodal discourse analysis, in: Hyland, K., Paltridge, B., Wong, L. (Eds.), Bloomsbury Companion to Discourse Analysis, second ed. London: Bloomsbury.

O'Halloran, K. L., Wignell, P., Tan, S., 2021. Big Data and managing multimodal complexity, in: Caldwell, D., Martin, J. R., Knox, J. (Eds.), Developing Theory: A Handbook in Appliable Linguistics and Semiotics. Bloomsbury, London.

Peeples, L., 2020. What the data say about police brutality and racial bias — and which reforms might work. Nature, 583, 22-24.

Pfeffer, J., Mayer, K., Morstatter, F., 2018. Tampering with Twitter's sample API. EPJ Data Science, 7(50).

Reilly, M., Thompson, S., 2017. Reverse image lookup: assessing digital library users and reuses. Journal of Web Librarianship, 11(1), 56–68.

Seng, K. P., Ang, L.-M., Liew, A. W.-C., Gao, J. (Eds.), 2019. Multimodal Analytics for Next-Generation Big Data Technologies and Applications. Springer, New York.

Sethy, A., Ramabhadran, B., 2008. bag-of-word normalized n-gram models. INTERSPEECH-2008: 9th Annual Conference of the International Speech Communication Association, Brisbane, Australia, 22-26 September 2008. https://www.isca-speech.org/archive/interspeech_2008/i08_1594.html

Stevanovic, D., Vlajic, N., An, A., 2013. Detection of malicious and non-malicious website visitors using unsupervised neural network learning. Applied Soft Computing, 13(1), 698–708.

Stöckl, H., Caple, H., Pflaeging, J. (Eds.), 2020. Shifts Towards Image-Centric Practices in the Contemporary Multimodal Practices. Routledge, New York.

Tan, S., O'Halloran, K. L., Wignell, P., Chai, K., Lange, R., 2018. A multimodal mixed methods approach for examining recontextualisation patterns of violent extremist images in online media. Discourse, Context and Media, 21, 18-35. doi:https://doi.org/10.1016/j.dcm.2017.11.004

Tan, S., O'Halloran, K. L., Wignell, P., 2020. Multimodality, in: Fina, A. D., Georgakopoulou, A. (Eds.), Handbook of Discourse Studies. Cambridge University Press, Cambridge, pp. 263-281.

Wang, G., Li, C., Wang, W., Zhang, Y., Shen, D., Zhang, X., Henao, R., Carin, L., 2018. Joint embedding of words and labels for text classification. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, July 15 - 20, 2018. https://arxiv.org/abs/1805.04174

Wevers, M., Smits, T., 2020. The visual digital turn: using neural networks to study historical images. Digital Scholarship in the Humanities, 35(1), 194-207.

Zhuang, W., Jiang, Q., Xiong, T., 2012. An intelligent anti-phishing strategy model for phishing website detection, in: 2012 32nd International Conference on Distributed Computing Systems Workshops, IEEE, pp. 51–56. https://ieeexplore.ieee.org/document/6258133

---

[1] https://en.wikipedia.org/wiki/George_Floyd

[2] https://sproutsocial.com/features/social-media-analytics/

[3] https://www.clarifai.com/

[4] https://cloud.google.com/

[5] https://en.wikipedia.org/wiki/Google_Compute_Engine

[6] https://www.splunk.com/

[7] https://docs.clarifai.com/

[8] https://metro.co.uk/video/michael-gove-says-occasions-driven-test-eyesight-2179184/

[9] https://en.wikipedia.org/wiki/Black_Lives_Matter

[10] https://www.independent.co.uk/voices/joe-biden-you-aint-black-republicans-democrats-donald-trump-a9533701.html