

Employing machine learning for capturing COVID-19 consumer sentiments from six countries: a methodological illustration

COVID-19
consumer
sentiments

869

Bodo B. Schlegelmilch
WU Vienna University of Economics and Business, Vienna, Austria
Kirti Sharma
Management Development Institute, Gurugram, India, and
Sambbhav Garg
University of Petroleum and Energy Studies, Dehradun, India

Received 4 June 2021
Revised 17 August 2021
Accepted 12 November 2021

Abstract

Purpose – This paper aims to illustrate the scope and challenges of using computer-aided content analysis in international marketing with the aim to capture consumer sentiments about COVID-19 from multi-lingual tweets.

Design/methodology/approach – The study is based on some 35 million original COVID-19-related tweets. The study methodology illustrates the use of supervised machine learning and artificial neural network techniques to conduct extensive information extraction.

Findings – The authors identified more than two million tweets from six countries and categorized them into PESTEL (i.e. Political, Economic, Social, Technological, Environmental and Legal) dimensions. The extracted consumer sentiments and associated emotions show substantial differences across countries. Our analyses highlight opportunities and challenges inherent in using multi-lingual online sentiment analysis in international marketing. Based on these insights, several future research directions are proposed.

Originality/value – First, the authors contribute to methodology development in international marketing by providing a “use-case” for computer-aided text mining in a multi-lingual context. Second, the authors add to the knowledge on differences in COVID-19-related consumer sentiments in different countries. Third, the authors provide avenues for future research on the analysis of unstructured multi-media posts.

Keywords Multi-media posts, Computer-aided analysis, Text mining, Multi-lingual tweets, Consumer sentiments, Supervised machine learning, Artificial neural network, PESTEL, Cross-country comparisons

Paper type Research paper

Introduction

The COVID-19 pandemic has had a devastating impact, not only in terms of health and human suffering, but also in terms of business (Apedo-Amah *et al.*, 2020; Bagchi *et al.*, 2020). In most countries, shops were closed, supply chains were interrupted and scores of people lost their jobs. From many companies active in international marketing, the COVID-19 pandemic represents the ultimate nightmare scenario. While the ability of companies to respond to the pandemic differs as much as the ability and resolve of countries to fight the pandemic, virtually all international marketers have had to adapt their products and services to the changing business environment and the shifting expectations and priorities of their



© Bodo B. Schlegelmilch, Kirti Sharma and Sambbhav Garg. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

customers across the world. To this end, international marketers require a sound understanding of how the pandemic has affected people worldwide, and what fears, concerns, hopes, observations, expectations and opinions on the pandemic they express. These days, an important source for gaining such information is social media. This is where many consumers share their commentary, and this is where international market researchers can gain rich insights into their thinking.

However, analyzing such unstructured text is complex. Metaphors such as “hitting the nail on the head,” emoticons like smiley faces, acronyms including “lol” or sound-like interjections like “wow” are often difficult to interpret. To complicate things further, international marketing researchers must also deal with text in foreign languages emanating from different cultures. Cultures have different norms, express emotions differently (Tsai, 2007) and use culture-specific metaphors or idioms. They may even differ in their use of pronouns, such as a preference for “we” in collectivistic versus “I” in individualist cultures (Berger *et al.*, 2020). In short, the interpretation of text from different languages and cultures is challenging, full of ambiguities and requires personal expertise from researchers to get it right. As today’s big data environment, described by the 5 Vs – “volume,” “velocity,” “veracity,” “variety” and “value” – mostly consists of unstructured text (Gandomi and Haider, 2015), interpretation of such data demands the use of computer-aided analysis tools. No individual researcher would, for example, be able to analyze a few million tweets in an efficient and timely manner.

This paper attempts to illustrate to international marketers, how unstructured multi-lingual text related to the COVID-19 pandemic can be analyzed with publicly available tools, including supervised machine learning and artificial neural network (ANN) techniques. We provide a step-by-step account ranging from extensive data pre-processing to information extraction and analysis. Our research is based on some 35 million original COVID-19-related tweets. The aim was to identify tweets from six selected countries, namely, Brazil (BRA), Germany (DEU), Great Britain (GBR), India (IND), Italy (ITA) and the USA, and subsequently use these to distill, categorize and analyze key topics and emotions discussed during the COVID-19 pandemic. The selection of the specific countries was driven by their geographic, cultural and economic (healthcare spending) differences.

Taken collectively, the paper aims to make four contributions. First, in terms of theory, we advance the understanding of the role of multi-lingual consumer sentiments in the analysis of environmental context factors commonly used in drafting international marketing strategy. Moreover, to the best of our knowledge, this is the first paper that tracks five machine-extracted distinct emotions across different countries. This extends our understanding of the different emotional responses across cultures and, consequently, helps us to theorize about the role and intensities of emotions in different countries.

Second, in terms of methodology, we offer a “use-case” for state-of-the-art computer-aided text mining in a multi-lingual context. Using sentiments connected to the COVID-19 crisis as context, we demonstrate to international marketers how unstructured text from different countries can be analyzed and interpreted. We assume that most international market researchers are not developing their own algorithms for conducting the essential steps required in this analysis, such as the structuring of location data, translating foreign language sentiments into English or annotating consumer sentiments for further classification, and are illustrating how commercially available programs can be used in this process.

Third, in terms of managerial insights, we offer a detailed categorization of COVID-19-related consumer sentiments and associated emotions in different countries. This will offer international marketers some pointers on the substantial country-specific differences that exist in terms of such sentiments and emotions, and can serve as a first basis for adapting marketing measures to the COVID-19 crisis across countries. Such marketing measures can

reach from taking steps to reduce the human interaction in the sales and delivery of goods to drafting advertising with COVID-19-related contents.

Fourth, we discuss the challenges and limitations inherent in using computer-aided multi-lingual sentiment analysis in an international marketing context. While the applied analytical approaches are potentially very powerful and provide a glimpse into the future of international marketing research, there are still substantial limitations to overcome. We hope to contribute to a wider application of machine learning in international marketing by identifying key research areas that still need to be addressed before computer-aided multi-lingual sentiment analysis can become mainstream.

The paper is structured as follows. We first review the literature on online sentiment analysis. This provides the basis for understanding the methodology we are using. We then explain our rationale for using the PESTEL framework to classify topics extracted from consumer tweets in different countries and outline our research propositions. Next, we describe our methodology, including data collection, structuring of location data, conversion of foreign language tweets into English, classifying tweets into PESTEL dimensions and associating the topics with sentiments. The subsequent section presents the results, including the key COVID-19-related topics addressed in the tweets by PESTEL category and country, and the sentiments around the most frequently occurring topics across all countries. The paper closes with a discussion of the main opportunities and challenges inherent in using sentiment analysis for unstructured multi-lingual text and identifies avenues for future research.

Literature review and research proposition

Sentiment analysis and Twitter data

When studying social media content, a distinction must be made between structured and unstructured data. Structured data refer to numerical or temporal data, while unstructured data refer to natural language text, images, audio or video that cannot be directly used as input to a program that can describe or analyze the data. Given that some 80% of social media content consists of unstructured data (Gandomi and Haider, 2015; Schneider, 2016), approaches to analyzing such data, say in the form of natural language texts, are much sought after.

Advancements in techniques like part-of-speech (PoS) tagging, named entity recognition (NER), information extraction and sentiment analysis (Toutanova *et al.*, 2003; Finkel *et al.*, 2005; Devika *et al.*, 2016) have made applications of natural language processing (NLP) relatively conspicuous, and the analysis of unstructured textual data more accessible. Sentiment analysis has emerged as a means to systematically extract and classify opinions, attitudes, thoughts, judgments and emotions from user-generated Web content (Thet *et al.*, 2010; Yu and Hatzivassiloglou, 2003; Rambocas and Gama, 2013; Qazi *et al.*, 2017). Online sentiment analysis denotes different analytical approaches, including supervised and unsupervised machine learning methods as well as lexicon, keyword and concept-based approaches (Li *et al.*, 2019; Kumar and Garg, 2020). Automated sentiment analysis relies on algorithms to process text. Tang *et al.* (2014) used a dictionary of affective words from SentiStrength2, while Ludwig *et al.* (2013) used the linguistic inquiry and word count program (LIWC), which calculates the proportion of words in a text that matches the pre-defined dictionaries.

Studying the effect of positive, negative and neutral user-generated comments on sales, Sonnier *et al.* (2011) were among the first in the marketing literature to model the dynamic effects of online communication via sentiment analysis. Other prominent examples include contributions by Tirunillai and Tellis (2012), who based their research on aggregated user-generated content from multiple websites and shed light on the relationship between online

media commentary and stock market performance. [Netzer et al. \(2012\)](#) listened to consumers' real-time discussion in online forums, while [Ludwig et al. \(2013\)](#) employed text mining to extract the semantic content and linguistic style of book reviews on [Amazon.com](#) to study retailers' conversion rates. In a similar vein, [Homburg et al. \(2015\)](#) examined consumer reactions toward firms' participation in consumer-to-consumer online community conversations, and [Liang et al. \(2015\)](#) focused on mobile app reviews and conducted sentiment analysis on this textual data. Other noteworthy contributions include papers by [Tang et al. \(2014\)](#), [Schweidel and Moe \(2014\)](#), [Hennig-Thurau et al. \(2015\)](#), [Pathak and Pathak-Shelat \(2017\)](#), [Micu et al. \(2017\)](#), [Kauffmann et al. \(2020\)](#) and [Luo et al. \(2021\)](#). More recently, [Rambocas and Pacheco \(2018\)](#) offered a review of various marketing applications of sentiment analysis. While most marketing applications analyze user-generated content relating to the purchase of goods and services, other topics have also been examined, such as consumer boycotts ([Makarem and Jae, 2016](#)), prediction of election results ([Budiharto and Meiliana, 2018](#)) and gender bias in sentiment analysis ([Thelwall, 2018](#)).

Several sentiment analyses have used Twitter data for their research (e.g. [Ji et al., 2015](#); [Lee, 2013](#); [Lansdall-Welfare et al., 2016](#); [Ibrahim et al., 2017](#)) due to Twitter's policy of making tweets highly accessible for social media content analysis. Twitter data are a combination of text, emoticons and hashtags. Its length limit of 280 characters makes it well suited for sentiment analysis ([Ji et al., 2015](#)). In fact, other neuro-linguistic programming solutions are unlikely to work well with Twitter data ([Katz et al., 2015](#)).

While sentiment analysis in general, and the analysis of Twitter data in particular, have made remarkable inroads into marketing, to the best of our knowledge, such techniques have not yet been used in an international context. The likely reasons lie in the problems inherent in dealing with foreign language texts, namely, the translation of text into one common language, and the complexity in interpreting statements that are steeped in cultural meaning.

Comparison of consumers across countries and research propositions

International marketing researchers and practitioners know that consumer behavior, attitudes, interests, opinions and emotions differ considerably between countries. Much of international marketing focuses on balancing tensions between universalism and country-specific differences. Nearly all cross-country consumer differences are rooted in environmental factors. These include differences in culture, economic and geographical conditions, as well as political and legal factors. Various approaches to systematizing the analysis of environmental differences between countries have been proposed. One of the most widely used frameworks to analyze macro-environmental factors is the PESTEL analysis, which focuses on political, economic, social, technological, environmental and legal dimensions ([Gillespie, 2007](#); [Grant, 2019](#); [Wurthmann, 2020](#)). The analysis has been applied in a broad range of contexts (e.g. [Kolios and Read, 2013](#); [Song et al., 2017](#)), most recently also in connection with the COVID-19 pandemic ([Thakur, 2021](#)), and is part of the standard toolbox of most international marketing researchers.

We use PESTEL, as this framework captures a wide range of factors that may explain differences between consumers' commentaries on the COVID-19 pandemic. For example, political factors may influence consumers' trust in government measures to curb the spread of the virus (e.g. [Flinders, 2020](#)). Economic and social factors (e.g. [Chen et al., 2020a, b](#); [Saladino et al., 2020](#); [Severo et al., 2021](#)), which include crucial aspects like cultural norms and values (e.g. [Pogrebna and Kharlamov, 2020](#)), health consciousness (e.g. [Pu et al., 2020](#)) and education level (e.g. [Marinoni et al., 2020](#)), may drive the need to continue working, the ability to cope with the consequences of the pandemic and appreciation of the virus' severity. Technological factors (e.g. [Dwivedi et al., 2020](#); [Urbaczewski and Lee, 2020](#)) may drive belief in scientific solutions to fight the pandemic. Environmental factors (e.g. [Cheval et al., 2020](#);

Shakil *et al.*, 2020), including climate, could influence consumers' willingness to shield from the virus. Finally, legal factors (Pistor, 2020) may impinge on employment laws and job security.

Based on our observations on sentiment analysis and Twitter data on the one hand, and the likely impact of country-specific macro-environmental factors on consumer sentiments on the other hand, we set out to demonstrate the scope for using machine learning in capturing and interpreting multi-lingual tweets relating to the COVID-19 pandemic. More specifically, we propose that sentiments related to the COVID-19 pandemic are likely to differ across countries, particularly when these countries show substantial differences in national culture, economic development and healthcare provision. We expect this to manifest in country differences in pandemic-related topics and the intensity of emotions expressed in the respective tweets.

Methodology

Figure 1 illustrates the sequence of steps we used in our analyses, ranging from data collection to the determination of topics and emotions from tweets. As the technical discussions and results generated in this process are rather extensive, much had to be placed in two Web-Appendices. Web-Appendix-A contains all tables and figures we generated in the course of our analysis, while Web-Appendix-B includes additional methodological details. We therefore request interested readers to consult the two Web-Appendices in conjunction with the paper.

Data collection

We aimed to analyze data from the following six countries: Brazil (BRA),

Germany (DEU), Great Britain (GBR), India (IND), Italy (ITA) and the USA. Three factors primarily drove the choice of countries. We wanted to obtain first a geographic spread, and second, a culturally diverse sample. The Hofstede scores (Hofstede, 2011) indicate substantial differences between the selected countries. Third, we were interested in the potential impact of healthcare spending on consumer sentiments. Data from the Organisation for Economic Co-operation and Development (OECD) [1] illustrate the vast differences in healthcare spending, ranging from US\$10,586/head in the USA to US\$209/

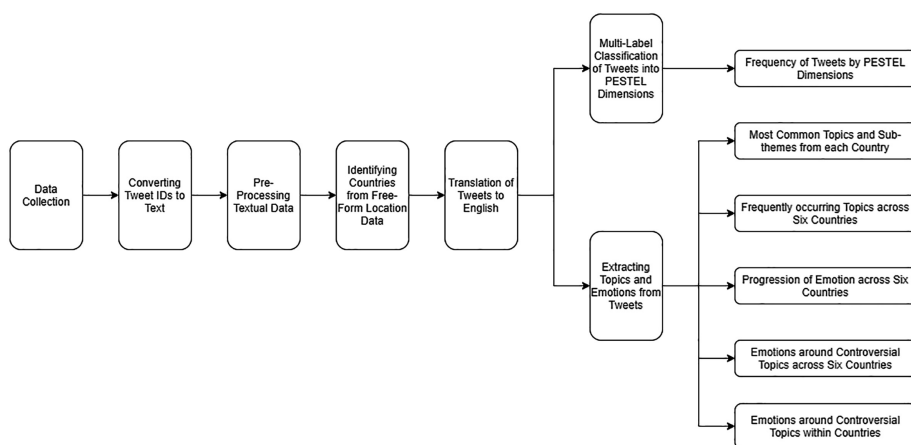


Figure 1.
Sequence of
analytical steps

head in India. As far as COVID-19 data are concerned, four of our six focal countries saw a first peak of the pandemic nearly simultaneously, while two saw the first peak afterward [2, 3, 4].

We used Twitter as a data source since it permits access to authentic consumer statements without engaging in costly surveys or qualitative data-gathering techniques. We collected about 35 million COVID-19-related tweets from an open-source GitHub repository (Chen *et al.*, 2020a, b) [5]. However, this database includes tweets from various geographic locations. Moreover, due to Twitter's terms of usage and data policy, the GitHub repository contains the data in the form of Tweet IDs instead of the complete tweet and its meta-data. Tweet IDs cannot be used to perform any textual analysis. To this end, we initially had to convert (hydrate) the raw data into a text format to be analyzed. Only after this initial data cleaning could we extract tweets from the six countries four weeks before the first peak in each country. We have chosen the Python (Van Rossum and Drake, 2009) programming language to process data at all steps.

Hydrating raw data

Table 1 shows an example of the resulting data hydration. To illustrate the process, the database authors have chosen to track tweets using keywords such as "Coronavirus," "Corona," "CDC," "Wuhan," "N95," "Epidemic," "outbreak," "covid19," "corona virus," "pandemic." The following is the list of Twitter accounts the database authors have chosen to track – "@CoronaVirusInfo," "@V2019 N," "@CDCemergency," "@CDCgov," "@WHO," "@HHSgov," "@NIAIDNews," "@drtedros" (The complete list can be found in the published GitHub repository).

Timestamp	Text	Free-form text location
Sun Mar 01 01:11:11 +0000 2020	No quiero ser mal pensada . . . pero, a los chinos se les escapo el virus . . . y ahoracomercializan un spray ðŸŒ°ðŸ™,ðŸŒ°	Santiago, Chile
Sun Mar 01 01:11:11 +0000 2020	RT @JuddLegum: There are only TWO coronavirus cases in Mexico There are 16 in Canada For Trump, every problem has the same solution	My house
Sun Mar 01 01:11:11 +0000 2020	RT @DarwinV_Quiroga: Cuando el Coronavirus estabaatacando a paÅses del primer mundo y de pronto llegaa Ecuador . . . #CoronavirusEc https://a !	Ecuador
Sun Mar 01 01:11:11 +0000 2020	RT @ABC: Top NIH official Anthony Fauci says coronavirus vaccine will likely take 12 to 18 months: "That means . . . the answer to containing iâ€!	Worcester MA
Sun Mar 01 01:11:11 +0000 2020	RT @chrishayes: Itâ€™s not coming from Mexico!!!!!!!	-
Sun Mar 01 01:11:11 +0000 2020	RT @intheMatrixxx: Who thinks he has it? Pope Francis sick a day after supporting coronavirus sufferers https://t.co/Sp17IXFNtA	Kawartha Lakes, Ontario
Sun Mar 01 01:11:11 +0000 2020	RT @BobRmhenry1: Trump sets new travel restrictions over coronavirus, considering southern border shutdown https://t.co/CNTsMUXNo6	Royal Oak, Michigan
Sun Mar 01 01:11:11 +0000 2020	RT @frontlinepbs: As 2047 â€ the moment that China takes full control of Hong Kong â€ approaches, the protesters FRONTLINE followed worry aboâ€!	-

Table 1.
Results of the data hydration

Data pre-processing

It is important to process every string of text to remove unwanted characters/words that could cause inconsistency in a model’s predictions. We adopted the following steps to remove noise from the location data and tweets: (1) Converting all text to lowercase, (2) removal of special characters, (3) removal of hyperlinks, (4) removal of Twitter mentions (“RT @user”), and (5) removal of stop words. The role of stop words during data pre-processing has been widely discussed (Keogh *et al.*, 2004; Xu and Wunsch, 2005; Rose *et al.*, 2010) and aims to reduce the complexity of the analysis and increase the efficiency of selecting useful terms (Sinka and Corne, 2005; Manco *et al.*, 2002).

Structuring of location data

To develop a machine learning model for classifying country locations based on free-form location fields present in tweets’ meta-data, we used HERE Technologies’ [6] geocoding service to label the data, since it would be inefficient to manually resolve the names of all countries, states, cities, towns and streets entered by a user when they first registered for Twitter. Hence, we labeled about 750,000 locations, sufficient to deploy a NLP to classify the remaining tweet locations in our dataset into their respective countries. Of the 750,000 labeled samples, we kept labeled locations from the 30 most frequently occurring countries to train a supervised machine learning model. We used the term frequency – inverse document frequency (TF-IDF) word-embedding model on the following machine learning algorithms: (1) linear regression, (2) support vector classifier (SVC), and (3) multinomial naïve Bayes (Multinomial NB). The NLP algorithms used are given in [Web-Appendix-A](#), with metrics depicting how each algorithm performed over our dataset (Kim and Gil, 2019; Vinodhini and Chandrasekaran, 2014).

We chose the SVC, as it returned the highest accuracy and F1-score. The model returns the probability of a string belonging to all the countries in our dataset. During the prediction, we set a minimum confidence level of 90% and kept only those countries that were relevant to this study. [Web-Appendix-A](#) shows the resulting country-wise volume of tweets, and [Table 2](#) provides a snapshot of the dataset with a column named “mapped.” This column shows from which country a user is tweeting from, as predicted with more than 90% confidence by the machine learning model.

Timestamp	Text	Free-form text location	Mapped Location
Mar 01	rt india temporarily suspended flights iran friday	anstrutherscotland	GBR
Mar 01	coronavirus county oof	sheher	GBR
Mar 01	rt feels abandoned amp forgotten prison iran infecting evin prison	kinross	GBR
Mar 01	rt announced baby engagement	cirencesterlondon	GBR
Mar 01	absolutely heartbreaking make sure trending woman needs taken home safety family immediately truly horrific	londonengland	GBR
Mar 01	rt bit luck boris announces yet another baby day pritipatel coronavirus floods would got	london	GBR
Mar 01	reached mumbai	hertfordshire	GBR
Mar 01	rt story every front page	nwengland	GBR
Mar 01	rt borisjohnson foreign secretary amp made false claim used justify nazanins imprisonment jo	crawleyenglanduk	GBR

Table 2.
Data enriched with
country associations

Conversion of foreign language tweets into English

For both GBR and the USA, the prevailing language used is English. For other countries, tweets had to be translated into English since the labeled data used to train the multi-label classification model in this study is English (Table 3). We translated non-English tweets using a software library in Python built to utilize Google Translate [7].

Following the described data labeling and cleaning procedures, we distilled 4.2 million tweets relating to the four weeks during which the COVID-19 pandemic peaked in the respective countries.

Classifying tweets into Political, Economic, Social, Technological, Environmental and Legal (PESTEL) dimensions

Most content we consume daily is in the form of text. Often such text needs to be classified, for example, to enable a news agency to recommend articles to enrich the user experience. One way is to manually classify and store each article inside folders. Another is to employ classical machine learning-based models such as those previously discussed. However, most classical machine learning-based models require in-depth feature engineering with human expert assistance to describe the texts' patterns. Moreover, classical machine learning-based models' performances increase with the data to a certain point, after which they cannot take complete advantage of large datasets due to their pre-defined features (Minaee et al., 2020).

Given the large number of tweets (approximately four million), it is nearly impossible to annotate those in the sample manually; hence, we used IBM Watson natural language understanding's (NLU) [8] application programming interface (API), which classifies tweets into over 500 predefined categories per request. For instance, if the pre-processed sentence coronavirus symptoms flu gathers is queried, the response is/health and fitness/disease/cold and flu. This is known as taxonomic hierarchy-based (THB) classification. We utilized the free-tier from IBM Watson NLU to classify over 600,000 tweets to create a labeled dataset for the multi-label text classification problem. We chose an average of eight THB classes for each PESTEL dimension, as mentioned in Table 4, and filtered out tweets having any other THB class. After the non-PESTEL tweets were filtered about 277,000 PESTEL tweets remained and were used to create the multi-label text classifier using a specific type of ANN called the recurrent neural network (RNN) [9]. The number of tweets remaining in each country after the PESTEL classification of tweets with 80% prediction confidence is given in Web-Appendix-A.

Determining topics and emotions from tweets

Similar to content classification, IBM Watson NLU offers services to enrich high-level topics present in text. For example, if a certain tweet is talking about the COVID-19 pandemic's effects on Brazil's economy, the service would return the following list – ["SARS,"

Country	% tweets in certain language	Tweets remaining*
Brazil	English: 54.03%, Portuguese: 31%, Spanish: 7%	151,554
Germany	German: 57.5%, English: 42.4%	96,563
Great Britain	English: 91.5%	481,707
India	English: 85.9%	242,475
Italy	Italian: 74.6%, English: 17%	150,353
USA	English: 81%	3,074,949

Table 3.
Volume of tweets in
different languages

Note(s): * after converting non-English tweets to English and removing textual noise

<i>Social</i>	<ul style="list-style-type: none"> • /hobbies and interests/games' • /art and entertainment/shows and events' • /art and entertainment/movies and tv' • /travel • /education/home schooling' • /education/school' • /family and parenting' • /food and drink/dining out' • /religion and spirituality' • /shopping' • /society/dating' • /society/gay life' • /society/welfare' • /society/racism' • /society/senior living' • /society/work' • /society/unrest and war' • /science/social science' • /sports' 	<ul style="list-style-type: none"> • /law, govt and politics/government' • /law, govt and politics/politics' • /law, govt and politics/legal issues' 	<ul style="list-style-type: none"> • /education' • /society/work' • /shopping/retail' • /business and industrial' • /careers/job search' • /careers/career planning' • /finance' 	
<i>Technological</i>	<ul style="list-style-type: none"> • /business and industrial/business software • /home and garden/appliances' • /automotive and vehicles' • /science/computer science' • /engineering' • /technology and computing' • /hobbies and interests/art and technology' 			
<i>Political</i>				
<i>Economic</i>				
<i>Legal</i>				
<i>Environmental</i>				

Table 4.
Categories as returned
by the IBM Watson
NLU API grouped into
PESTEL

“Pandemic,” “Brazil,” “Economics”]. Additionally, IBM Watson NLU detects five distinct emotions in textual documents: (1) Anger, (2) disgust, (3) fear, (4) joy, and (5) sadness. A numerical value is returned for each emotion based on the results of a machine learning algorithm. The higher the value returned for each emotion, the more likely the tweet expresses the particular emotion. Table 5 represents the topics and emotions collected and stored using the Python programming language for Brazil, returned by the IBM Watson NLU API. The “emotions” returned by IBM Watson are not to be confused with the complete spectrum of human emotions [10]. Neither does it play the role of psychological innuendo meant to quantify human emotions.

Results

The first steps of our analysis are structured around the PESTEL dimensions. More specifically, across all six countries, we initially analyze the inter-category frequency of PESTEL tweets, and then show a 30-day sentiment analysis averaged over each day per PESTEL dimension. Subsequently, our analysis becomes more fine-grained, and we present comparisons between topic frequencies and the most common underlying themes observed with the most frequent topics. Finally, our results focus on individual emotions, emotions pertaining to each country, emotions around the most-frequent topics occurring commonly across all countries and five key emotions across the most-frequent topics.

Each tweet can be associated with one or more dimensions in the PESTEL framework and can belong to multiple PESTEL dimensions, based on the six predictions returned by the RNN model. Figure 2 depicts a frequency-normalized inter-category trend between six dimensions over the PESTEL framework. We normalize the data by scaling the number of

Text	Concepts	Anger	Disgust	Fear	Joy	Sadness
ok show source deaths unduly covid man spent suffering hours without break didnt eat went bathroom sat hour evolve patients tweet beyond irresponsible covid becoming rare testing expanding asymptomatic people either ignorant . . .	["Brazil", "SARS coronavirus"]	0.30719	0.12082	0.24642	0.00926	0.66467
corona virus natural disasters case bubonic plague braineating amoeba incompetent government like bolsonaro toure museum fire political opponents hopefully dies corona virus pls go away george vlog floridaaaa coronavirus hand . . .	["Black Death", "Bubonic plague", "SARS coronavirus"]	0.31097	0.14952	0.14921	0.03287	0.67907
many great ideas thoughts school reopening latest nightly family neighbors country wear mask son palestinian woman infected climbed hospital room sit see mother every pride n absence cure exican scientists develop drug prevents . . .	["Brazil"]	0.03781	0.07432	0.21957	0.3585	0.47979

Table 5. Example of concepts and emotions as returned by the IBM Watson NLU API for BRA

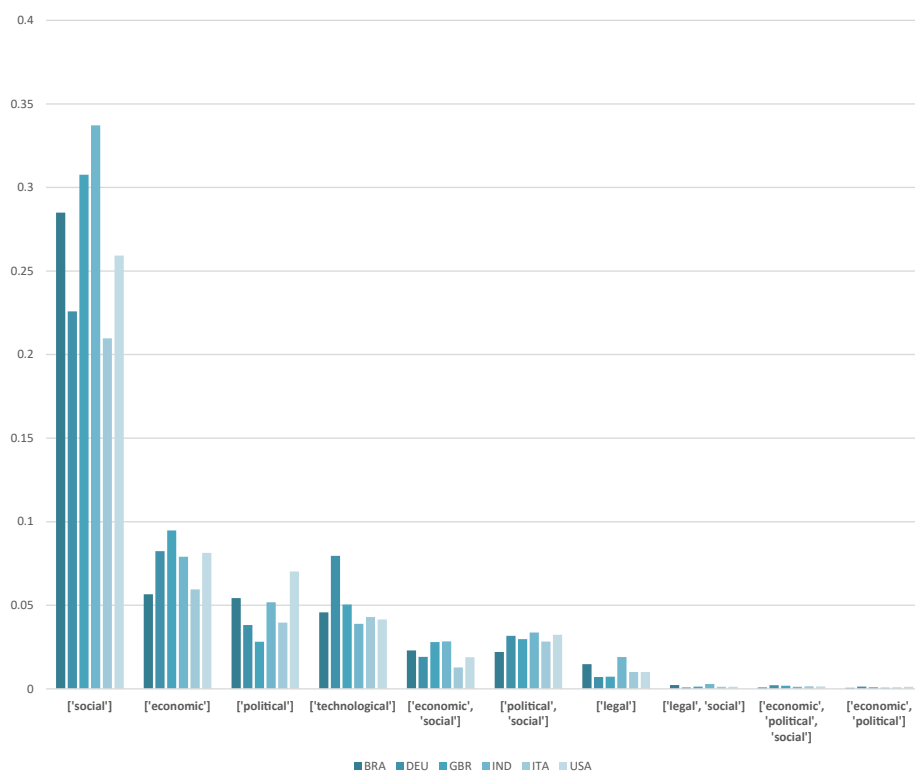


Figure 2.
Tweets by PESTEL
dimensions

tweets between zero and one, drawing comparisons between different countries, as the total number of tweets is different for each country. Overall, most tweets related to the social dimension of PESTEL, whereas comparatively few tweets related to the legal dimension. Within each PESTEL dimension, Twitter users in the six analyzed countries tweeted with different intensity. For example, Indian Twitter users tweeted most strongly about topics with a social dimension, British in the economic dimension, US Twitter users lead in political tweets, German Twitter users tweeted most strongly within the technological dimension and Indian Twitter users more frequently about legal topics. These country differences are not only statistically significant, but also quite substantial. For example, US Twitter users tweet more than twice as much about politics than British Twitter users. Similarly, German Twitter users tweet more than twice as much about technological topics than Indian Twitter users. Indian Twitter users tweeted twice as much about legal topics than all other analyzed countries except Brazil. The other dimensions shown in [Figure 2](#) are mixed PESTEL dimensions and capture tweets related to two or more of the PESTEL dimensions.

[Figure 3](#) represents a comparison of the consumer tweets in the 24 most frequent topics. The frequencies are normalized to the total number of tweets in each country to draw comparisons between them, since the total numbers of tweets are different. Unsurprisingly, the most frequently tweeted topics in nearly all countries related to “Severe Acute Respiratory Syndrome,” the exception being those from UK Twitter users, who tweeted most frequently about “English-Language Films.” Overall, films featured highly in tweets during the first wave of the pandemic, presumably reflecting how many Twitter users

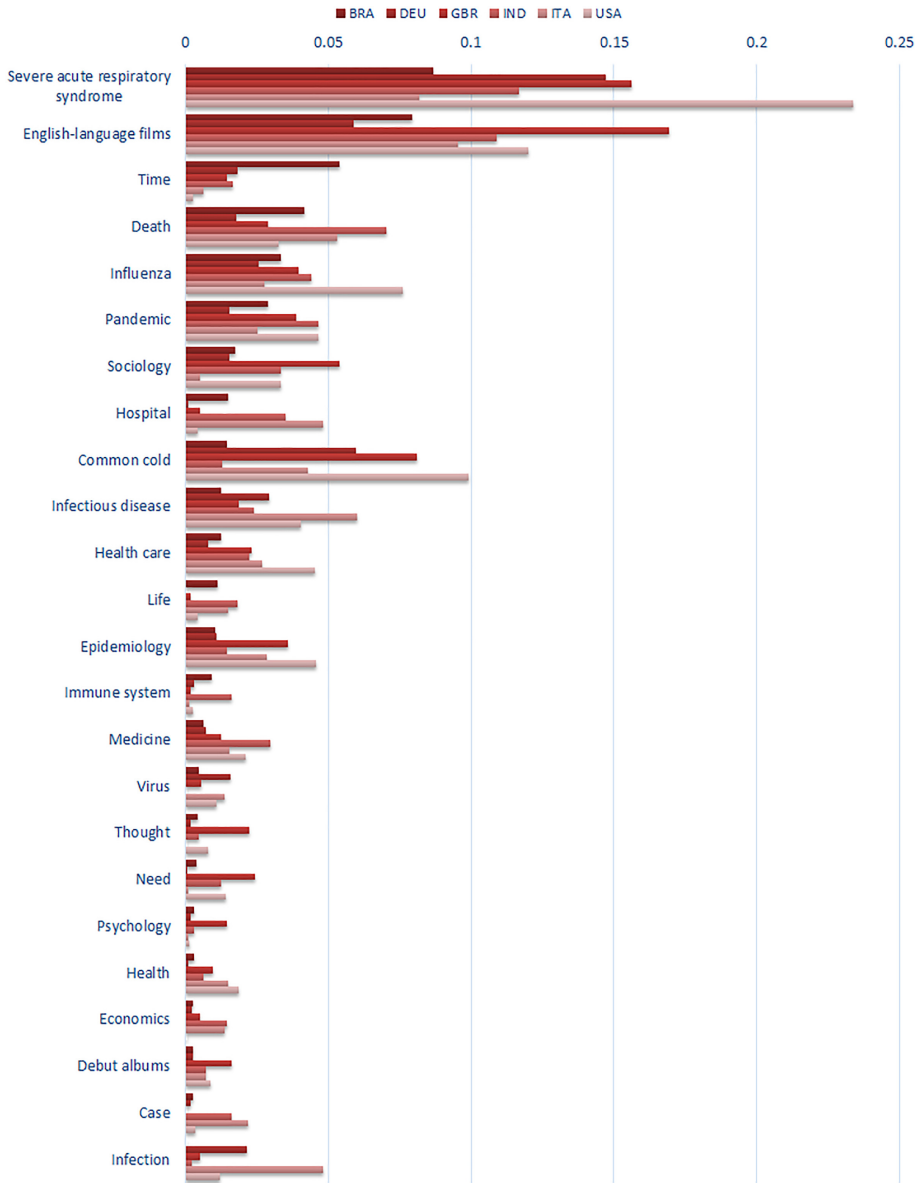


Figure 3.
Comparing frequency
of concepts across six
countries

occupied themselves by watching films during lockdown. Once again, there are substantial differences in the prominence of tweeted topics across the six analyzed countries, with the ratio between the countries tweeting least and most about a given topic sometimes as high as 1:2.9 (S.A.R.S.; ITA:USA). Noteworthy differences occur, for example, on the topics of “Hospital” and “Health Care,” where German Twitter users tweeted much less in comparison to other countries. The topic “Hospital” concerned Italian Twitter users the most, coinciding with Italy’s severe hospital crises during the initial phase of the pandemic

(Sanfelici, 2020). Germany, in contrast, has the highest number of intensive care beds per 1,000 people in Europe (Farr, 2020).

US Twitter users tweeted most heavily about “Health Care” (22 times more than German Twitter users). A possible explanation might lie in the comparatively large number of uninsured or underinsured US citizens (Cohen *et al.*, 2019) [11]. Noteworthy are the large differences in tweets mentioning “Common Cold.” The high proportion of such tweets from the USA and the UK may reflect the initial propensity of the political leaderships in both countries to downplay the virus as no more dangerous than a cold (Brooks, 2020).

Next, we drilled one level deeper to understand the most frequent underlying themes. Given this journal’s length restrictions, Table 6 only provides an excerpt of this analysis [12], showing the most frequent topics and themes in Columns (2)–(7) for each country. At this level of aggregation, it is more difficult to identify country-specific patterns; nevertheless, some patterns emerged. For example, all mentions of “Mask” (save for a small number in Italy) appear in tweets from Brazil and India. A similar country-specific concentration occurs when looking at the terms “Paper” or “Toilet Paper.” German, and to a lesser extent US and British Twitter users, were obsessed with toilet paper shortages, while the topic was not mentioned in Brazil, Italy or India. That said, there is also a range of common topics, such as “Illness,” “Vaccine or Vaccination,” Public Health, “Health Insurance” and “Health Economics,” which occur across all countries. In a few instances, differences between countries can be explained by different linguistic preferences, for example, in using “Aging,” “Gerontology” or “Old Age” and using “Swine” versus “Pig.” Taken collectively, a closer look at the individual terms reveals a mix between some pronounced idiosyncrasies and commonalities in the topics across all six countries.

Subsequently, we determined how a particular country performed over the five distinct emotions mentioned earlier. Since we collected the data points from the four weeks up to COVID-19 peaks, we have visualized a temporal representation of emotions for each country. To illustrate the analysis, we focus on the development of “Sadness” (Figure 4) and “Fear” (Figure 5) [13] When looking at the development of “Sadness,” we see that the level remains

Topic	BRA	DEU	GBR	IND	ITA	USA
Severe acute respiratory syndrome	Mask	Animal virology	Hygiene	Government	Chinese language	Swine influenza
	Public health	Paper	Government	Mask	Animal virology	Avian influenza
	Smallpox	Toilet paper	Animal virology	Swine influenza	Public health	Public health
	High school	Federalism	Swine influenza	Smallpox	Swine influenza	Paper
	Swine influenza	World population	Influenza pandemic	Public health	Italian language	Influenza pandemic
	Avian influenza	Hygiene	Paper	Sovereign state	Influenza pandemic	Animal virology
	World population	German language	Public health	Avian influenza	Avian influenza	Animal virology
	Civil and political rights	Swine influenza	Avian influenza	Vaccination		Vaccination
	Education	Federal government	Smallpox	2009 flu pandemic	Chinese language	

Table 6.
Excerpt of most common topics and sub-themes observed in each country

IMR
40,5

882

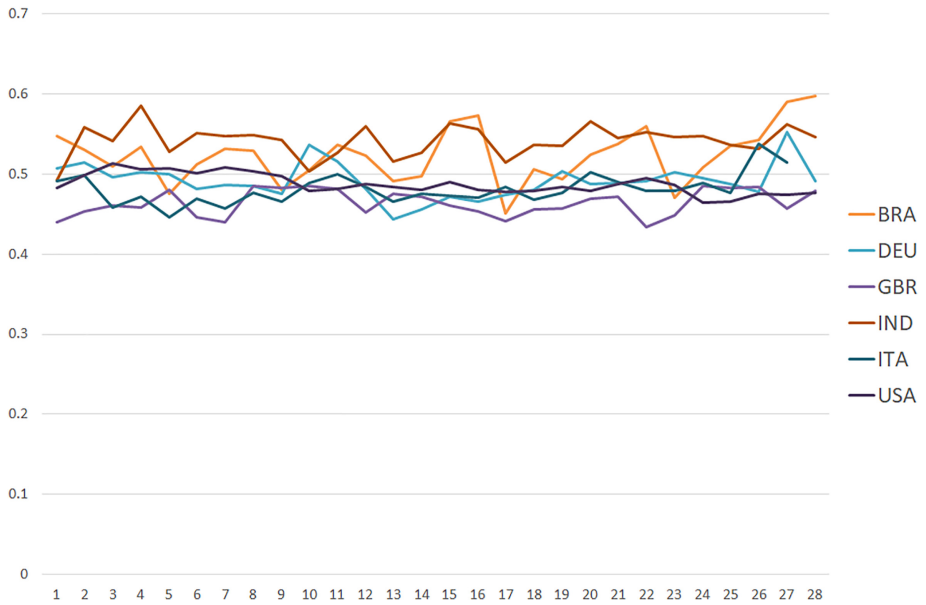


Figure 4.
Progression of sadness
in six countries

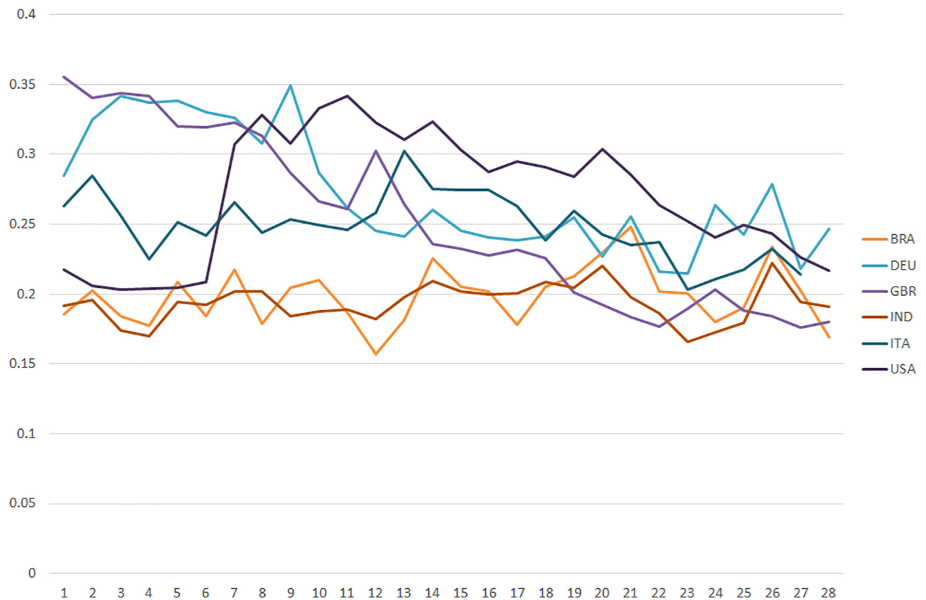


Figure 5.
Progression of fear in
six countries

relatively high across all six countries during this time period. By contrast, looking at “Fear,” we observe much more cross-country variation and a lower level of intensity. While the level of “Fear” remains relatively stable in India and Brazil, we observed a sharp rise starting on Day 6 in the USA, followed by a gradual decline as we approached the end of the observation period. In Britain, “Fear” started at a comparatively high level at the beginning of the

observation period and then declined to the second-lowest level of all six countries at the end of the first peak of the pandemic.

Shifting our vantage point to comparing emotions within each country, we use India and the USA to illustrate our analysis (Figures 6 and 7) [14]. As in the other analyzed countries, “Sadness” is the most pronounced emotion in India. The other four emotions develop at a notably lower level and remain in a relatively narrow range of fluctuation. By contrast, in the USA, “Sadness,” “Anger” and “Disgust” also fluctuate in narrow ranges, but “Fear” and “Joy” show substantial variations during the period leading up to the first peak of the pandemic. The sudden spike in “Fear” and, conversely, the sudden decline of “Joy” at the beginning of March 2020, is quite remarkable. Our results coincide with a Gallup poll conducted between March 21 and April 5, which reported that daily stress and worry plagued 60% of American adults; stress was five times and worry four times higher than during the 2008 recession (Witters and Harter, 2020).

Next, we extract the six most frequent topics that were common to all six countries: (1) “Severe Acute Respiratory Syndrome”; (2) “English-language Films”; (3) “Infectious Disease”; (4) “Influenza”; (5) “Pandemic”; and (6) “Death.” This enables us to compare emotions expressed by each country for that particular topic. Figure 8 illustrates the result when focusing on “Infectious Disease.” Most striking, once again, is the dominance of “Sadness,” followed by “Fear,” the second most dominant emotion in all analyzed countries, albeit comparatively lower in Brazil and Italy. While the emotional responses to the topic “Infectious Disease” vary between countries, the overall pattern appears rather similar. This also holds for the emotional responses to the five other topics singled out for this analysis [15].

Last, we aggregated the topics and emotions from the scores returned by the IBM Watson NLU API by country. Analyzing the five distinct emotions relating to the most frequently

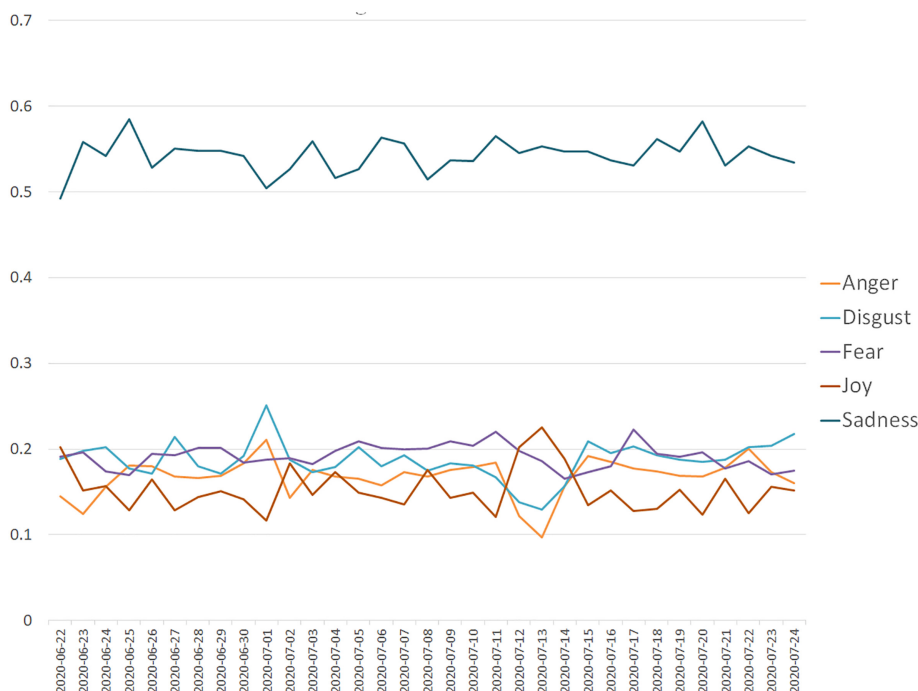


Figure 6.
Progression of emotion
in India

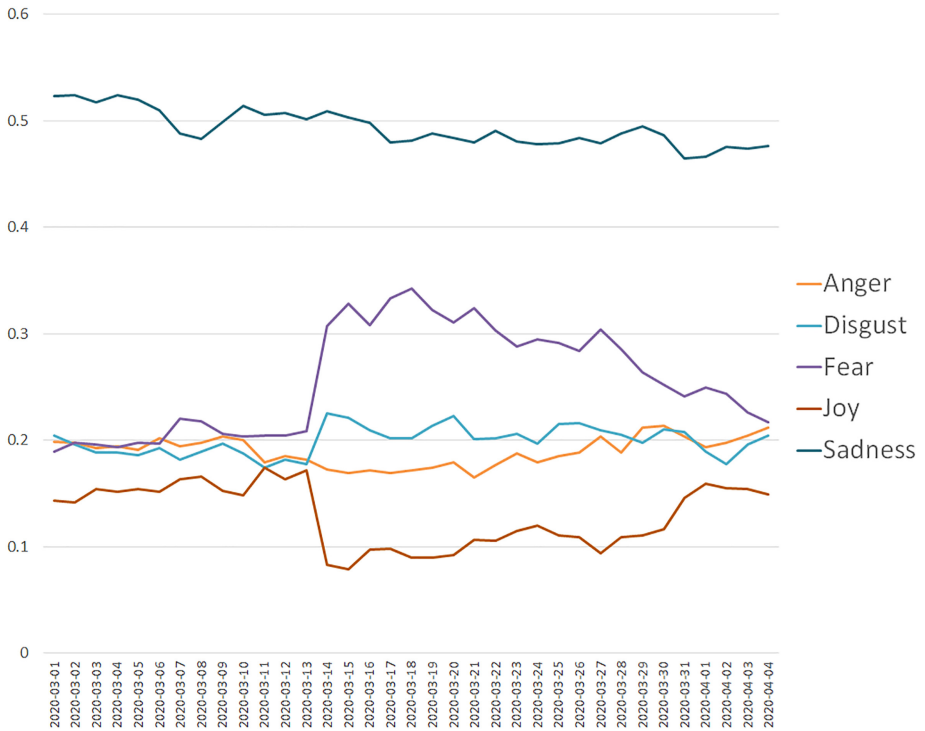


Figure 7.
Progression of emotion
in the USA

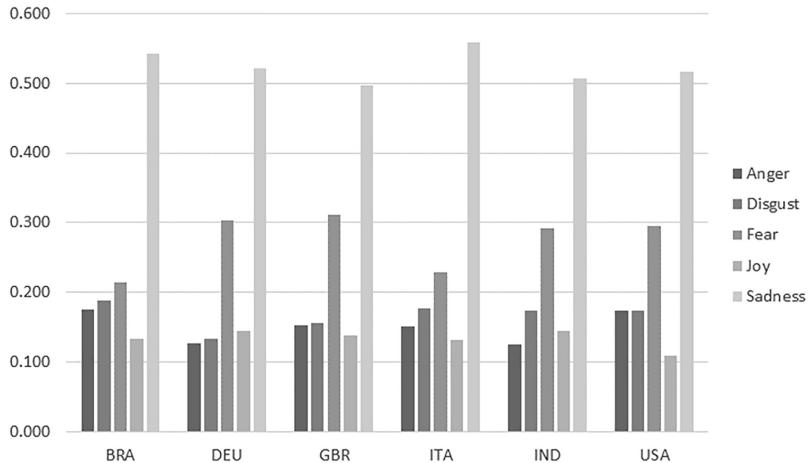


Figure 8.
Emotions around the
topic "Infectious
Disease" across six
countries

occurring topics, we can visualize the results in bar charts. To illustrate, [Figure 9](#) shows the results for Germany [16]. Once again, we can see the dominance of "Sadness," and to a somewhat lesser extent "Fear" as the most frequent topics.

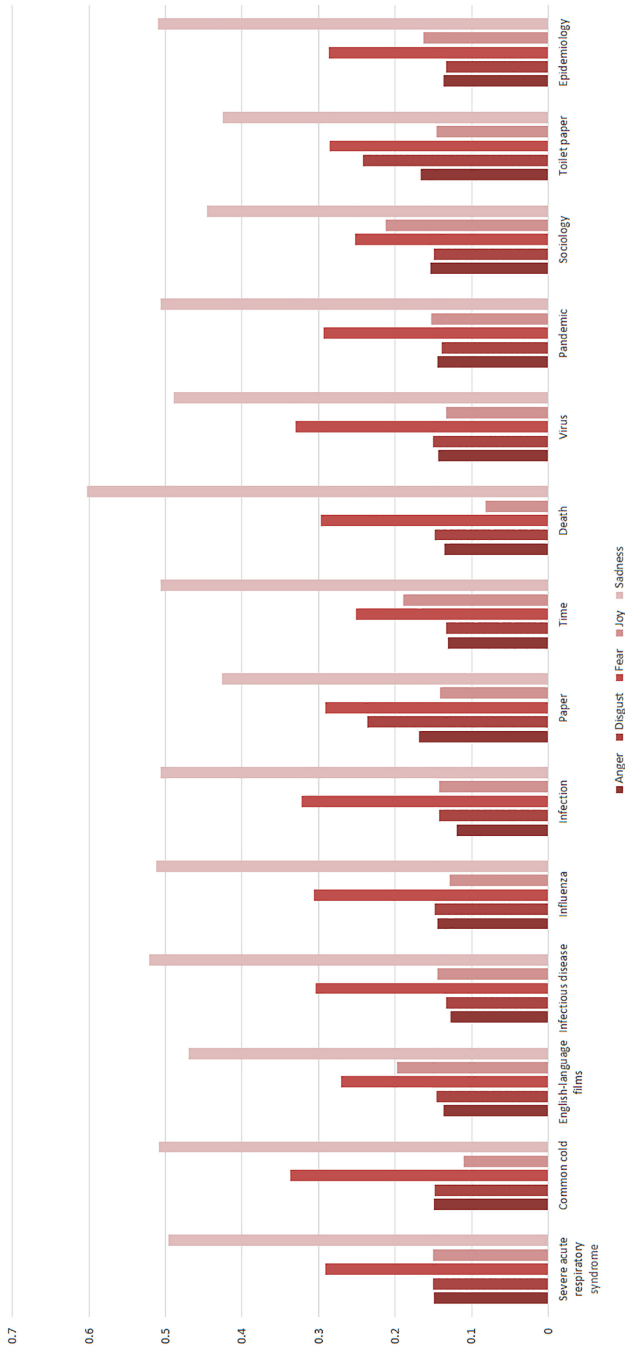


Figure 9.
Graphical
representation of
emotions around top
concepts in Germany

Various additional analyses would be possible based on the original tweets with the topics returned by the IBM Watson NLU API. When filtering tweets, for example, by the topic “Influenza,” in India, the following keywords are found: “advice,” “stay healthy,” “still stay home,” “wash hands frequently,” “keep distance,” “pandemic rages.” Using the same “Influenza” filter in Germany, the keywords identified were: “prevent spread virus,” “do not know enough,” “do not vaccine cure.” However, such an analysis would go beyond the scope of this paper due to the sheer size of the resulting tables.

Discussion and suggestions for future research

Using tweets on COVID-19 as a “use case,” this paper illustrates the process of extracting and analyzing online sentiments from multi-lingual tweets with computer-aided content analysis. The described approach clearly shows it is possible to analyze vast numbers of social media postings to generate detailed insights into consumer sentiments across different countries. That said, our analyses also demonstrate the challenges inherent in using computer-aided content analysis in an international marketing context.

An initial hurdle lies in the market penetration differences of social media platforms in different countries. Looking specifically at Twitter in our focal countries, as of July 2020, the USA reported the highest number of Twitter users with 62.55 million. This was followed by India, Brazil and Great Britain with 17 million, 15.7 million and 15.25 million active users, respectively. Germany and Italy have the lowest number of users, with 4.2 million and 3.2 million, respectively (We Are Social, 2019). Consequently, there may be differences in Twitter users’ socio-demographic composition and even large samples are unlikely to represent the overall populations in each country. This may result in biases and should be considered when interpreting the findings. Future research should, therefore, provide insights into the composition of users of social media platforms in different countries to get a feel for the representativeness of the analyzed data.

A second challenge consisted of the considerable data loss at each stage of the data extraction and pre-processing. Starting with a database of some 35 million COVID-19-related tweets from all over the world, the number of tweets in our dataset reduced down to five million, because we chose to analyze only six countries. In addition, we imposed a constraint of 90% prediction confidence on the location classification algorithm to maintain quality result. Future research may choose to broaden the horizon by choosing to analyze more countries and possibly employ a neural network approach for the same. Translation of non-English tweets to English with the Google Translate Python wrapper library further reduced the data to 4.2 million tweets. Finally, classifying the remaining tweets into PESTEL dimensions, using an 80% prediction confidence as a cut-off rate, reduced those available for analysis to 2.2 million. Thus, future research should put a strong focus on how data loss can be reduced without compromising the quality of the remaining data.

Third, in an international marketing context, the quality of automatic translations remains a concern. Given the number of words with multiple meanings, something may be “lost in translation.” Take the German word “verlegen,” which could mean “not able to find something,” “to publish something” or “to feel embarrassed.” While considerable advances have been made in recent years, capturing culture-specific idiosyncrasies – particularly important when dealing with differences in expressing emotions – merits more research. While considerable advances have been made in recent years, the area of automatic translation deserves additional research efforts.

A fourth issue is related to categorizing tweets into PESTEL dimensions. To achieve this, we utilized the free-tier from IBM Watson NLU to classify over 600,000 tweets to create a labeled dataset for the multi-label text classification. Tweets enriched using IBM Watson NLU

API for topics and emotions were requested in groups of ten to reduce the number of requests made to the IBM server. This was necessary to remain within the API service limits for individual IBM Watson accounts. Combining tweets into ten for this section hardly vitiates the findings of this paper, as the topics returned from a particular tweet are limited to a considerably larger number, and the emotions returned from a particular tweet usually convey the overall emotion. Thus, we trust that our cost-efficient approach has only a negligible effect on the quality of the analysis. Nevertheless, future researchers could achieve an even finer-grained analysis if they are willing to spend money on commercially available services.

Having outlined some of the challenges inherent in the computer-aided analysis of multi-lingual tweets, we now focus on the opportunities offered by this approach and look at some of the findings of our COVID-19 “use case.” To the best of our knowledge, this paper is the first contribution categorizing some two million tweets into PESTEL dimensions using an RNN. This provides a good understanding of the relative emphasis that Twitter users place on the different PESTEL dimensions, and enables us to attach five distinct emotions to the classified text elements, a considerable advance over the usual “positive,” “neutral” and “negative” classification offered by most sentiment analyses. The ability to track these emotions between and within countries provides policymakers with a measurement tool that captures social media reactions to pandemic and policy announcements. In fact, the analyses can be quite detailed, in that it is possible to compare the development of emotions for broad topics and detailed themes within or across countries, and trace these emotions over time. This can offer policymakers and international marketing managers pointers for communication campaigns. Policymakers, for example, may more effectively address privacy concerns related to the use of corona tracking apps, uncertainty about the use of face masks, or skepticism toward COVID-19 vaccinations, as they learn which concerns are particularly pertinent and how emotionally these topics are discussed. Similarly, international marketers can use the analytical approach to fine-tune their advertising campaigns. We can already observe how a number of brands have incorporated COVID-19-related themes in their advertising. Examples from different countries range from Burger King’s “Pay Cut Whopper,” H&M’s “Wearable Love” to Pizza Hut’s “Have fun staying at home” ([Ads of the World, n.d.](#)) to Nike’s “Play for the World” with sad eerie shots of empty basketball courts and stadiums ([Nielsen, 2020](#)).

Conclusion

Our paper has sought to develop a use case for a computer-aided sentiment analysis of multi-lingual tweets. We demonstrated that it is possible to conduct a PESTEL analysis based on social media posts from different countries. This enabled us to identify and categorize insightful cross-country differences in consumer interests and concerns about the COVID-19 crisis. Additionally, we showed it is possible to track consumer emotions relating to (broader) topics or (more specific) themes over time and conduct between- and within-country comparisons.

While the main contribution of the paper is methodological, in that it focuses on the scope for using computer-aided sentiment analysis in a multi-lingual international marketing context, our findings also contribute to theory development, provide insights for policymakers and international marketing managers, and point to some promising future research avenues.

As to theory development, we show that multi-lingual consumer sentiments can be meaningfully categorized into PESTEL dimensions, which, in turn, inform the drafting of international marketing strategy. This extends our understanding of the different roles of macro-factors and emotional responses across cultures and, consequently, helps us to theorize about the role and intensities of emotions in different countries.

In terms of managerial implications, we offer international policymakers and marketers some insights into the substantial country specific differences in consumer concerns and emotions regarding the COVID-19 pandemic. To this end, our findings can serve as a basis for adapting marketing measures relating to the COVID-19 crisis across countries. Marketers in different parts of the world have already adapted their marketing to the COVID-19 crisis. Advertising, for example, has been created with COVID-19 related content, which addresses different concerns (e.g. income shortfall) and emotions (e.g. sadness) in different countries. Our analyses can give international marketers data on the specific concerns and emotions that are relevant in different country markets.

Finally, with reference to future research directions, our paper also demonstrated some of the key challenges in using computer-aided online sentiment analysis for multi-lingual social media data from different countries. Concerns relating to data representativeness, the large loss of data during pre-processing and the ability of automated translation programs to capture linguistic subtleties across different cultures all lead to a rich agenda for future research.

Taken collectively, our paper illustrates the scope for using online sentiment analysis for vast numbers of multi-media posts across different countries. However, it also highlights the need for caution when using online sentiment analysis in international marketing and shows that considerable research efforts are still needed to move this promising analytical toolbox into the mainstream of international marketing research.

Notes

1. <https://data.oecd.org/healthres/health-spending.htm>.
2. [New COVID-19 Cases Worldwide - Johns Hopkins Coronavirus Resource Center \(jhu.edu\)](#).
3. [Mortality Analyses – Johns Hopkins Coronavirus Resource Center \(jhu.edu\)](#).
4. Please see [Web-Appendix-A](#) for details on the COVID-19 impact in each of the six countries.
5. The data collected from this GitHub repository contained only COVID-19 tweets ([Chen et al., 2020a, b](#)). Note that a “free of cost research” API by Twitter was released after we had already collected, cleaned and processed the data.
6. [Guide - HERE Geocoding & Search API - HERE Developer](#).
7. <https://pypi.org/project/googletrans/>.
8. [Watson Natural Language Understanding | IBM](#).
9. The specifics of the multi-label text classification model can be found in [Web-Appendix-B](#).
10. We recognize that psychologists are strict about how and when to use the term “emotion.” In the case of text, it is often not apparent if the expressed sentiments are moods, feelings or emotions. Consequently, the term “affect” is often used in text mining, and the task of extracting it is labeled as “affect detection.” However, in this context, we maintain the term “emotion,” as used by IBM Watson.
11. [FastStats - Health Insurance Coverage \(cdc.gov\)](#).
12. The complete table of all 24 topics and themes across the six analyzed countries is available in [Web-Appendix-A](#).
13. The graphical representation of the other emotions is available in [Web-Appendix-A](#).
14. Please see [Web-Appendix-A](#) for the results of the other four countries.
15. Please see [Web-Appendix-A](#) for the emotional responses to the five other topics.
16. Please see [Web-Appendix-A](#) for the other five countries.

References

- Ads of the World (n.d), "COVID-19 ads", from COVID-19 Ads | Ads of the World™, available at: https://www.adsoftheworld.com/collection/covid19_ads (accessed 8 August 2021).
- Apedo-Amah, M.C., Avdiu, B., Cirera, X., Cruz, M., Davies, E., Grover, A., Iacovone, L., Kilinc, U., Medvedev, D., Maduko, F.O., Poupakis, S., Torres, J. and Tran, T.T. (2020), "Unmasking the impact of COVID-19 on business: from level evidence from across the world", World Bank Group, Policy Research Working Paper 9434, from World Bank Document (accessed 3 August 2021).
- Bagchi, B., Chatterjee, S., Ghosh, R. and Dandapat, D. (2020), "Impact of COVID-19 on global economy", in *Coronavirus Outbreak and the Great Lockdown*, Springer, Singapore, pp. 15-26.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W.W., Netzer, O. and Schweidel, D.A. (2020), "Uniting the tribes: using text for marketing insight", *Journal of Marketing*, Vol. 84 No. 1, pp. 1-25.
- Brooks, B. (2020), "Like the flu? Trump's coronavirus messaging confuses public, pandemic researchers say", *Thomson Reuters*, available at: <https://www.reuters.com/article/us-health-coronavirus-mixed-messages-idUSKBN2102GY> (accessed 12 April 2021).
- Budiharto, W. and Meiliana, M. (2018), "Prediction and analysis of Indonesia presidential election from Twitter using sentiment analysis", *Journal of Big Data*, Vol. 5 No. 1, pp. 1-10.
- Chen, E., Lerman, K. and Ferrara, E. (2020a), "Tracking social media discourse about the COVID-19 pandemic: development of a public Coronavirus Twitter data set", *JMIR Public Health and Surveillance*, Vol. 6 No. 2, p. e19273.
- Chen, S., Igan, D., Pierri, N. and Presbitero, A.F. (2020b), "Tracking the economic impact of COVID-19 and mitigation policies in Europe and the United States", working paper 20, 125, IMF, Washington, 10 July.
- Cheval, S., Mihai Adamescu, C., Georgiadis, T., Herrnegger, M., Piticar, A. and Legates, D.R. (2020), "Observed and potential impacts of the COVID-19 pandemic on the environment", *International Journal of Environmental Research and Public Health*, Vol. 17 No. 11, p. 4140.
- Cohen, R.A., Cha, A.E., Martinez, M.E. and Terlizzi, E.P. (2019), "National health interview survey early release program", *National Health Center for Health Statistics*, available at: <https://www.cdc.gov/nchs/data/nhis/earlyrelease/insur202009-508.pdf>.
- Devika, M., Sunitha, C. and Ganesh, A. (2016), "Sentiment analysis: a comparative study on different approaches", *Procedia Computer Science*, Vol. 87, pp. 44-49.
- Dwivedi, Y.K., Hughes, D.L., Coombs, C., Constantiou, I., Duan, Y., Edwards, J.S., Gupta, B., Lal, B., Misra, S., Prashant, P. and Raman, R. (2020), "Impact of COVID-19 pandemic on information management research and practice: transforming education, work and life", *International Journal of Information Management*, Vol. 55 No. 102211, pp. 1-20.
- Farr, C. (2020), "Germany's coronavirus response is a master class in science communication", available at: <https://www.cnbc.com/2020/07/21/germanys-coronavirus-response-masterful-science-communication.html> (accessed 12 April 2021).
- Finkel, J., Grenager, T. and Manning, C. (2005), "Incorporating non-local information into information extraction systems by Gibbs Sampling", Paper Presented at The Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL '05), Ann Arbor, Michigan, 25-30 June, pp. 363-370.
- Flinders, M. (2020), "Democracy and the politics of coronavirus: trust, blame and understanding", *Parliamentary Affairs*, available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7337828/> (accessed 12 April 2021).
- Gandomi, A. and Haider, M. (2015), "Beyond the hype: big data concepts, methods, and analytics", *International Journal of Information Management*, Vol. 35 No. 2, pp. 137-144.
- Gillespie, A. (2007), "PESTEL analysis of the macro-environment", *Foundations of Economics*, Oxford University Press, Oxford.
- Grant, R. (2019), *Contemporary Strategy Analysis*, 10th ed., John Wiley & Sons, New York, NY.

- Hennig-Thurau, T., Wiertz, C. and Feldhaus, F. (2015), "Does Twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies", *Journal of the Academy of Marketing Science*, Vol. 43 No. 3, pp. 375-394.
- Hofstede, G. (2011), "Dimensionalizing cultures: the Hofstede model in context", *Online Readings in Psychology and Culture*, Vol. 2 No. 1, pp. 1-26.
- Homburg, C., Ehm, L. and Artz, M. (2015), "Measuring and managing consumer sentiment in an online community environment", *Journal of Marketing Research*, Vol. 52 No. 5, pp. 629-641.
- Ibrahim, N., Wang, X. and Bourne, H. (2017), "Exploring the effect of user engagement in online brand communities: evidence from Twitter", *Computers in Human Behavior*, Vol. 72, pp. 321-338.
- Ji, X., Chun, S.A., Wei, Z. and Geller, J. (2015), "Twitter sentiment classification for measuring public health concerns", *Social Network Analysis and Mining*, Vol. 5 No. 13, pp. 1-25.
- Katz, G., Ofek, N. and Shapira, B. (2015), "ConSent: context-based sentiment analysis", *Knowledge-Based Systems*, Vol. 84, pp. 162-178.
- Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R. and Mora, H. (2020), "A framework for big data analytics in commercial social networks: a case study on sentiment analysis and fake review detection for marketing decision-making", *Industrial Marketing Management*, Vol. 90, pp. 523-537.
- Keogh, E., Lonardi, S. and Ratanamahatana, C. (2004), "Towards parameter-free data mining", *Paper Presented at The Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data*, Seattle, Washington, 22-25 August, pp. 206-215.
- Kim, S. and Gil, J. (2019), "Research paper classification systems based on TF-IDF and LDA schemes", *Human-centric Computing and Information Sciences*, Vol. 9 No. 3, pp. 1-21.
- Kolios, A. and Read, G. (2013), "A political, economic, social, technology, legal and environmental (PESTLE) approach for risk identification of the tidal industry in the United Kingdom", *Energies*, Vol. 6 No. 10, pp. 5023-5045.
- Kumar, A. and Garg, G. (2020), "Systematic literature review on context-based sentiment analysis in social multimedia", *Multimedia Tools and Applications*, Vol. 79 Nos 21-22, pp. 15349-15380.
- Lansdall-Welfare, T., Dzogang, F. and Cristianini, N. (2016), "Change-Point analysis of the public mood in UK Twitter during the Brexit referendum", *Paper Presented at The 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW 2016)*, 12-15 December, Barcelona, Spain, pp. 434-439, doi: [10.1109/ICDMW.2016.0068](https://doi.org/10.1109/ICDMW.2016.0068).
- Lee, J. (2013), "Validity of consumer-based physical activity monitors and calibration of smartphone for prediction of physical activity energy expenditure", *Graduate Theses and Dissertations*, Vol. 13480, Iowa State University, available at: <https://lib.dr.iastate.edu/etd/13480/> (accessed 16 November 2020).
- Li, Z., Fan, Y., Jiang, B., Lei, T. and Liu, W. (2019), "A survey on sentiment analysis and opinion mining for social multimedia", *Multimedia Tools and Applications*, Vol. 78 No. 6, pp. 6939-6967.
- Liang, T., Li, X., Yang, C. and Wang, M. (2015), "What in consumer reviews affects the sales of mobile apps: a multifacet sentiment analysis approach", *International Journal of Electronic Commerce*, Vol. 20 No. 2, pp. 236-260.
- Ludwig, S., De Ruyter, K., Friedman, M., Brüggem, E.C., Wetzels, M. and Pfann, G. (2013), "More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates", *Journal of Marketing*, Vol. 77 No. 1, pp. 87-103.
- Luo, J., Huang, S. and Wang, R. (2021), "A fine-grained sentiment analysis of online guest reviews of economy hotels in China", *Journal of Hospitality Marketing and Management*, Vol. 30 No. 1, pp. 71-95.
- Makarem, S. and Jae, H. (2016), "Consumer boycott behavior: an exploratory analysis of Twitter feeds", *Journal of Consumer Affairs*, Vol. 50 No. 1, pp. 193-223.

- Manco, G., Masciari, E. and Tagarelli, A. (2002), "A framework for adaptive mail classification", Technical report, ICAR-CNR, presented at The Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence, 4-6 November, Washington, DC, pp. 387-439.
- Marinoni, G., Land, H. and Jensen, T. (2020), "The impact of Covid-19 on higher education around the world", IAU Global Survey Report, available at: https://www.iau-aiu.net/IMG/pdf/iau_covid19_and_he_survey_report_final_may_2020.pdf (accessed 16 November 2020).
- Micu, A., Micu, A.E., Geru, M. and Lixandriou, R.C. (2017), "Analyzing user sentiment in social media: implications for online marketing strategy", *Psychology and Marketing*, Vol. 34 No. 12, pp. 1094-1100.
- Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M. and Gao, J. (2020), "Deep learning based text classification: a comprehensive review", available at: <https://arxiv.org/pdf/2004.03705.pdf> (accessed 12 January 2021).
- Netzer, O., Feldman, R., Goldenberg, J. and Fresko, M. (2012), "Mine your own business: market-structure surveillance through text mining", *Marketing Science*, Vol. 31 No. 3, pp. 521-543.
- Nielsen, S. (2020), "5 COVID-19 ads that stood out from the crowd", *CKP Communications Group*, from 5 COVID-19 ads that stood out from the crowd — CKP Group (theckpgroup.com) (accessed 8 August 2021).
- Pathak, X. and Pathak-Shelat, M. (2017), "Sentiment analysis of virtual brand communities for effective tribal marketing", *Journal of Research in Interactive Marketing*, Vol. 11 No. 1, pp. 16-38.
- Pistor, K. (2020), *Law in the Time of COVID-19*, Columbia Law School, New York, NY, available at: <https://scholarship.law.columbia.edu/books/240/> (accessed 12 January 2021).
- Pogrebna, G. and Kharlamov, A. (2020), "The impact of cross cultural differences in hand washing patterns on the COVID-19 outbreak magnitude", *Regulation and Governance*, Advance Online Publication, London, Vol. 10.
- Pu, B., Zhang, L., Tang, Z. and Qiu, Y. (2020), "The relationship between health consciousness and home-based exercise in China during the COVID-19 pandemic", *International Journal of Environmental Research and Public Health*, Vol. 17 No. 16, pp. 1-18, doi: [10.3390/ijerph17165693](https://doi.org/10.3390/ijerph17165693).
- Qazi, A., Raj, R.G., Hardaker, G. and Standing, C. (2017), "A systematic literature review on opinion types and sentiment analysis techniques", *Internet Research*, Vol. 27 No. 3, pp. 608-630.
- Rambocas, M. and Gama, J. (2013), "Marketing research: the role of sentiment analysis (No. 489)", Universidade do Porto, Faculdade de Economia do Porto.
- Rambocas, M. and Pacheco, B. (2018), "Online sentiment analysis in marketing research: a review", *Journal of Research in Interactive Marketing*, Vol. 12 No. 2, pp. 146-163.
- Rose, S., Engel, D., Cramer, N. and Cowley, W. (2010), "Automatic keyword extraction from individual documents", in Berry, M.W. and Kogan, J. (Eds), *Text Mining: Applications and Theory*, John Wiley & Sons, Hoboken, NJ, pp. 1-20.
- Saladino, V., Algeri, D. and Auriemma, V. (2020), "The psychological and social impact of Covid-19: new perspectives of well-being", *Frontiers in Psychology*, Vol. 11, doi: [10.3389/fpsyg.2020.577684](https://doi.org/10.3389/fpsyg.2020.577684) (accessed 11 January 2021).
- Sanfelici, M. (2020), "The Italian response to the COVID-19 crisis: lessons learned and future direction in social development", *The International Journal of Community and Social Development*, Vol. 2 No. 2, pp. 191-210.
- Schneider, C. (2016), "The biggest data challenges that you might not even know you have", *IBM*, available at: <https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know/> (accessed 13 April 2021).
- Schweidel, D. and Moe, W. (2014), "Listening in on social media: a joint model of sentiment and venue format choice", *Journal of Marketing Research*, Vol. 51 No. 4, pp. 387-402.
- Severo, E., De Guimarães, J. and Dellarmelin, M. (2021), "Impact of the COVID-19 pandemic on environmental awareness, sustainable consumption and social responsibility: evidence

- from generations in Brazil and Portugal”, *Journal of Cleaner Production*, Vol. 286, 124947, doi: [10.1016/j.jclepro.2020.124947](https://doi.org/10.1016/j.jclepro.2020.124947).
- Shakil, M.H., Munim, Z.H., Tasnia, M. and Sarowar, S. (2020), “COVID-19 and the environment: a critical review and research agenda”, *Science of The Total Environment*, Vol. 745, 141022, doi: [10.1016/j.scitotenv.2020.141022](https://doi.org/10.1016/j.scitotenv.2020.141022).
- Sinka, M.P. and Corne, D.W. (2005), “The BankSearch web document dataset: investigating unsupervised clustering and category similarity”, *Journal of Network and Computer Applications*, Vol. 28 No. 2, pp. 129-146.
- Song, J., Sun, Y. and Jin, L. (2017), “PESTEL analysis of the development of the waste-to-energy incineration industry in China”, *Renewable and Sustainable Energy Reviews*, Vol. 80, pp. 276-289.
- Sonnier, G., McAlister, L. and Rutz, O. (2011), “A dynamic model of the effect of online communications on firm sales”, *Marketing Science*, Vol. 30 No. 4, pp. 702-716.
- Tang, T., Fang, E. and Wang, F. (2014), “Is neutral really neutral? The effects of neutral user-generated content on product sales”, *Journal of Marketing*, Vol. 78 No. 4, pp. 41-58.
- Thakur, V. (2021), “Framework for PESTEL dimensions of sustainable healthcare waste management: learnings from COVID-19 outbreak”, *Journal of Cleaner Production*, Vol. 287, 125562.
- Thelwall, M. (2018), “Gender bias in sentiment analysis”, *Online Information Review*, Vol. 42 No. 1, pp. 45-57.
- Thet, T.T., Na, J. and Khoo, C. (2010), “Aspect-based sentiment analysis of movie reviews on discussion boards”, *Journal of Information Science*, Vol. 36 No. 6, pp. 823-848.
- Tirumillai, S. and Tellis, G. (2012), “Does chatter really matter? Dynamics of user-generated content and stock performance”, *Marketing Science*, Vol. 31 No. 2, pp. 198-215.
- Toutanova, K., Klein, D., Manning, C.D. and Singer, Y. (2003), “Feature-rich part-of-speech tagging with a cyclic dependency network”, Paper Presented at The Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Edmonton, Canada, 27 May-1 June, Vol. 1, pp. 173-180.
- Tsai, J. (2007), “Ideal Affect: cultural causes and behavioral consequences”, *Perspectives on Psychological Science*, Vol. 2 No. 3, pp. 242-259.
- Urbaczewski, A. and Lee, Y. (2020), “Information Technology and the pandemic: a preliminary multinational analysis of the impact of mobile tracking technology on the COVID-19 contagion control”, *European Journal of Information Systems*, Vol. 29 No. 4, pp. 405-414.
- Van Rossum, G. and Drake, F. (2009), *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA.
- Vinodhini, G. and Chandrasekaran, R. (2014), “Opinion mining using principal component analysis based ensemble model for e-commerce application”, *CSI Transactions on ICT*, Vol. 2 No. 3, pp. 169-179.
- We Are Social (2019), “Digital in 2019: global internet use”, available at: <https://wearesocial.com/uk/blog/2019/01/digital-in-2019-global-internet-use-accelerates/Accelerates>. <https://wearesocial.com/uk/blog/2019/01/digital-in-2019-global-internet-use-accelerates/>.
- Witters, D. and Harter, J. (2020), “Worry and stress fuel record drop in U.S. life satisfaction”, *Gallup.com*, available at: <https://news.gallup.com/poll/310250/worry-stress-fuel-record-drop-life-satisfaction.aspx>.
- Wurthmann, K. (2020), “The essential mix: six tools for strategy-making in the next decade”, *Journal of Business Strategy*, Vol. 41 No. 1, pp. 38-49.
- Xu, R. and Wunsch, D. II (2005), “Survey of clustering algorithms”, *IEEE Transactions on Neural Networks*, Vol. 16 No. 3, pp. 645-678.
- Yu, H. and Hatzivassiloglou, V. (2003), “Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences”, Paper Presented at The Proceedings of the Conference on Empirical Methods in Natural Language Processing, Sapporo, Japan, 11-12 July, pp. 129-136.

Appendixes

The Appendix files are available online for this article.

About the authors

Bodo B. Schlegelmilch is Professor and Chair of the Institute for International Marketing Management at WU Vienna University of Economics and Business, Austria. Bodo B. Schlegelmilch is the corresponding author and can be contacted at: bodo.schlegelmilch@wu.ac.at

Kirti Sharma is an Assistant Professor of Marketing and works at the Management Development Institute, Gurgaon, India.

Sambbhav Garg graduated in Computer Science and Engineering with a specialization in business analytics and optimization from the University of Petroleum and Energy Studies (UPES) in Dehradun, India.

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com