COVID-19 Impact Sentiment Analysis on a Topic-based Level

Mustapha Hankar[1,*], Marouane Birjali[1], Anas El-Ansari[2]
and Abderrahim Beni-Hssane[1]

[1]*LAROSERI Laboratory, Computer Science Department, University of Chouaib Doukkali, Faculty of Sciences, El Jadida, Morocco*
[2]*MASI Laboratory, Computer Science Department, FPN, Mohammed First University, Nador, Morocco*
E-mail: hankar.mustapha@gmail.com; marouane.birjali@gmail.com; anas.elansari@gmail.com; abenihssane@yahoo.fr
[*]Corresponding Author

Abstract

Last December 2019, health officials in Wuhan, a province from China, identified a novel coronavirus called SARS-CoV-2 causing pneumonia. In March 2020, World Health Organization (WHO) declared COVID-19 disease being a pandemic. During quarantine periods, people all over the globe were living under severe and overwhelming circumstances and expressing feelings of loneliness, dread, and anxiety. The pandemic has had a significant impact on the labor markets. As a result, several employees have lost their jobs while others are in grave danger to lose their positions the next day. In this paper, we developed a hybrid approach integrating sentiment analysis combined with topic modeling to analyze the impact of the COVID-19 pandemic on Moroccan citizens. The data used in this study includes comments collected from a well-known news website in Morocco called Hespress. Our approach

follows a two-step process. In the first step, we implement a topic modeling method to analyze and extract topics from Arabic comments, and in the second step, we perform topic-based sentiment analysis to classify people's feedback on extracted topics. The final results revealed that the expressed sentiments regarding all the topics are highly negative.

1 Introduction

Last December 2019, the novel coronavirus was discovered in Wuhan, a province in China calling health officials to trace the possible sources of the SARS-CoV-2 virus. Back then, they thought that its origin was a seafood market located in Wuhan [1]. As the disease expands its infection base exponentially, investigations are still conducted by researchers to reveal how coronavirus spreads. The main transmission mode [2] is via direct contact with droplets of an infected person coughing, sneezing, or touching contaminated surfaces. Months later, this epidemic is declared to be a pandemic by the world health organization (WHO), and many countries started to take harsh decisions to quarantine people at home as the only way to suppress the virus from spreading. Conducted studies [3, 4] revealed that protective measures during quarantine such as lockdowns, travel restrictions, and economic shutdowns have led many people to suffer from psychological disorders because of being isolated or socially distanced from each other, and also threatened of losing their jobs. The impacts of the lockdown include post-traumatic stress disorder symptoms, confusion, anger, long quarantine suppress symptoms, infection fears, frustration, boredom, lack of resources and supplies, and financial loss.

Sentiment analysis is very used in social media, microblogs, personalized systems [5, 6], and forums to retrieve and analyze textual data that holds an opinion of the user. Sentiment analysis enters the field of natural language processing (NLP), and it is widely applied in various domains such as marketing services to retrieve customers' opinions about a certain product, and social media to polarize opinions about political topics or public figures. Sentiment analysis methods are viewed at three main levels: document level, sentence level, and aspect or topic level [7]. This study focuses on performing sentiment analysis of the third type, namely the topic level.

Thus, a topic modeling model, namely latent Dirichlet allocation (LDA), is applied to extract topics from the corpus of collected documents. Afterwards, a topic-based sentiment classification is performed to classify comments into positive, neutral, and negative.

In this paper, we developed a hybrid approach that combines both topic modeling and sentiment analysis to analyze the impact of COVID-19 on Moroccan people during periods of quarantine. In this hybrid approach, we apply a topic modeling-based sentiment analysis to mine people's opinions on several extracted topics related to the COVID-19 pandemic. The system, in the first phase, extracts the topics from the dataset using LDA, and then a topic-based sentiment analysis classification is performed using a pre-trained model, namely CaMeL, in the second phase. The dataset provided for the study contains comments collected from Hespress[1] written in modern standard Arabic (MSA) and Moroccan dialect (MD). The system integrates many techniques and models to perform the topic-based sentiment classification on the collected comments. A topic modeling method is applied for topic extraction, a word embedding model is built to compute word similarities in the vector space [8], and a pre-trained sentiment classification model is implemented to classify subjective opinions about extracted topics related to the COVID-19 pandemic. For example, we want to know how COVID-19 affects various domains such as people's mental health, the state economy, the labour market, support plan launched by the government, or the impact of preventive measures on citizens' normal life, etc.

The remainder of this paper is organized as follows: we cite and discuss some related works to this study in the second section. The third section is dedicated to describing and exploring the dataset. Cleaning and pre-processing tasks are performed in the fourth section, with mentioning some pre-treatment issues in the Arabic language. In Sections 5 and 6, we cite the various methods used in the process of our work, such as LDA for topic modeling, word2vec word embedding model for similarity computing, and CaMeL pre-trained analyzer for sentiment analysis. In Section 7, we describe our hybrid approach to perform a topic-level sentiment analysis on COVID-19 comments. In Section 8, we present our findings for discussion and further perspectives. In the last section, we summarize our work and discuss some future work.

---

[1]https://www.hespress.com/

2 Related Work

In recent years, researchers have shown an important interest in natural language processing (NLP) in order to advance processing methods for a better understanding of human languages [9, 10]. However, many challenges are still facing other researchers in Arabic NLP due to specific characteristics of the language. In this section, we will reference and discuss some related works to our study.

Melville et al. [11] built a lexicon-based model for sentiment analysis trained on labeled documents. They combined lexicon and machine learning (Naïve Bayes) approaches to classify texts into positive, neutral, and negative sentiments.

Abdul-Mageed et al. [12], proposed a system (SAMAR) for Subjectivity and Sentiment Analysis (SSA) to classify the text into objective or subjective before identifying its polarity. The dataset used for this study was collected from social media websites, chat apps, Twitter, web forums, and Wikipedia Talk Pages. The system uses the SVMlight algorithm for the classification task. Their experiments faced some difficult and complex issues related to the characteristics of the Arabic language.

Hu et al. [13] used an unsupervised sentiment analysis method based on emotional signals. These signals were divided into two categories: emotion correlation and emotion indication. They conducted their approach on two Twitter datasets. The results proved the efficiency of their proposed model and the importance of the roles of different emoticons in sentiment analysis as well. In [14], Dasgupta et al. used a clustering technique, to group texts is into sentiment dimension. This happens by providing user feedback in the spectral clustering process in an interactive manner. Each dimension of spectral clustering contains features, which are considered sentiment-oriented topics. To determine the most important dimensions, they needed human interaction. The proposed framework needs a massive amount of manual annotation for documents to identify important dimensions.

In [15], Zarra et al. proposed a semi-supervised approach for colloquial Maghrebi Arabic which combines topic extraction and sentiment analysis in one model that assigns a polarity for each extracted topic. They used a supervised approach to classify sentiments, and an unsupervised approach to extract topics from a dataset of Facebook comments. Shelke et al. [16] proposed an approach for aspect-oriented and domain-independent sentiment analysis applied on product reviews. They implemented the SentiWordNet lexicon to detect the polarity of the review based on its identified features. Li et al. [17] developed a client perception-based product feature mining

technique that uses sentiment orientation. The feature and sentiment orientation of products are sorted by the weights of user interest to help customers access review orientation more effectively.

Akhtar et al. [18] proposed an aspect-specific framework for sentiment analysis that combines an aspect extraction method and a sentiment classification model that implements a particle swarm optimization (PSO) algorithm. The system works in a two-step process, aspect term extraction, and sentiment classification. They used three models for sentiment classification, namely, Maximum Entropy (ME), Conditional Random Field (CRF), and Support Vector Machine (SVM). Experiments of the system on two different kinds of domains showed high effectiveness.

Piryani et al. [19] proposed a linguistic rule-based aspect-level sentiment analysis approach for movie reviews. Their approach extracts the aspects from movie reviews, finds the opinion about the extracted aspect, and computes the polarity of the opinion using linguistic rules. According to the authors, this method showed good accuracy and may be integrated into an opinion profiling system. In [20], S. Asharaf et al. suggested an aspect-level sentiment analysis method applied on product reviews. They used LDA to extract aspect words and then mapped with various aspects of an entity to perform aspect-level sentiment analysis. Experiments on real-world datasets showed promising results compared to the state-of-the-art sentiment analysis methods.

In our previous work [21], we utilized a baseline bag of words (BOW) model to represent the documents as vectors. However, the model has many limitations such as the order of the words is ignored in the document representation and the semantic aspect of the text is undermined [22]. In this paper, we overcame this problem by reinforcing the proposed system with a word embedding model to enhance the topic extraction task from the corpus using word similarities [23]. We used principal component analysis (PCA) for dimensionality reduction to project the model in a two-dimensional space to identify topic words. More comments were added to the previous dataset to empower our system and to cover actual aspects related to COVID-19 pandemic.

3 Data Description

The dataset provided for this study contains collected comments from Hespress (online local news platform) written in both MSA and MD. The size of the previous dataset is increased to more than 21000 comments. We coded python crawler using selenium module to extract all the comments of opinion

articles and news articles related to the COVID-19 pandemic in various domains such as health, economy, politics, society, vaccine, etc. The Arabic comments were filtered from other comments written in other languages like English, French, or Spanish using a python function to detect the language of the comment. The columns names of the dataset are: the date when the comment was published, the user's name, and the comment text.

4 Data Cleaning and Pre-processing

Arabic language characteristics have made processing its text in the web more challenging wherein researchers must deal with several difficulties related to the nature of the language such as ambiguity, diglossia, and many other difficulties related to the morphology of Arabic [24]. Another challenge is that the web Arabic text is a mixture of classical Arabic (CA), modern standard Arabic (MSA), Moroccan dialect (MD), and may also contain Arabizi words written in Latin letters, notations, orthographic features like repeated letters, spelling mistakes, slang words, ironic expressions, idiomatic expressions [25]. Arabic language has different structure which makes the processing task more difficult than other languages. For example, Arabic is a right-to-left language with 28 letters including three vowels and special diacritics called harakat (حركات) like fatha (فتحة), sukun (سكون), hamza (همزة), tanwin (تنوين), shadda (شدة)…etc. Diacritics can change the semantics and the syntax of a word, making the meaning of the words in the text more confusing. The forms of letters alter according on which position they appear in a word [24]. Hence, cleaning a text written in Arabic includes removing punctuations, links, emojis, numbers, repeated letters, diacritics, and words that are not written in Arabic letters and stop words. For example, given an Arabic phrase like: "إِنَّ الْقُرَّأَءَ يَقْرَؤُوْنَ الْقُرْآنَ قِرَأَةً جَمِيْلَـــــةً". After cleaning this sentence, we will have an output sentence as follows: ان القراء يقرؤون القران قراءة جميلة. The next task to come after cleaning is to preprocess the output comments. This includes tokenization which means splitting the text into formal words called tokens, and stemming is performed on these tokens by eliminating both suffixes and prefixes to convert them into the base form of the terms called stems. The Table 1 below shows some examples of text comments before and after cleaning and preprocessing.

Table 1   Raw comments versus preprocessed comments

Comment	Pre-processed comment
الا بذكر الله تطمئن القلوب عزوزوا ايمانكم وفروا الى الله وقراءة القرآن فوالله لا علاج نفسي اقوى من هذا فجددوا ايمانكم تفوزوا بخير الدارين	ذكر, الله, تطمئن, قلوب, عزوزوا, ايمانكم, وفروا, الله, قراءة, قرآن, والله, علاج, نفسي, اقوى, جددوا, ايمانكم, تفوزوا, خير, دارين
و هدوك اصحاب الرميد... واصحاب الرميد المنتهية صلاحيته و القطاع الغير المهيكل وهدوك لي غلطو او غلطو فيهم و و و و و كاين منهم مشدش حتى الدفعة الاولى\n...زيد و	هدوك, اصحاب, رميد, اصحاب, رميد, منتهية, صلاحيته, قطاع, غير, مهيكل, هدوك, غلطو, فيهم, زيد, كاين, منهم, مشدش, دفعة
تعليق الدراسة حتى اخر الشهر ستساهم في وقاية الطلبة من وباء كورونا. والتقلص في انتشاره السريع	تعليق, دراسة, اخر, شهر, تساهم, وقاية, طلبة, وباء, كورونا, تقلص, انتشار, سريع
السيد الوزير هل انت واعي، جواز التلقيح هو الوسيلة الوحيدة للحماية من كورونا ، اه دبا الفيروس الا لقا عندي الجواز ميقسنيش اوا بلاش من اللقاح فرقوا علينا غير الجواز	سيد, وزير, انت, واعي, جواز, تلقيح, وسيلة, وحيدة, حماية, كورونا, دبا, فيروس, لقا, عندي, جواز, ميقسنيش, لقاح, فرق,جواز
المغرب أخذ احتياجاته منذ مدة ، ألم ينشئ صندوقا خاصا بالكوارث وتمت الزيادة في كل التأمينات؟ أم أن تلك المساهمات ذهبت لجهات أخرى بلا رجعة ؟	احتياج, مدة, ينشئ, صندوق, خاص, كوارث, تمت, زيادة, تأمين, مساهم, ذهبت , جهات, اخرى, رجعة

5 LDA for Topic Extraction

Topic modeling is widely used in the fields of text mining and information retrieval to extract the hidden topics (or items) in a collection of documents. In a process of topic modeling, we intend to discover the hidden topics (aspects) in the corpus. Each document from the corpus is considered as a set of various topics, and each topic is represented as a set of various words. For example, suppose we have a two-topic model of news articles. The distribution of topics in document 1 might be given as 90% topic A, 5% topic B, and 5% topic C, and the distribution of topics in document 2 is given as 30% topic A, 60% topic B, and 10% topic C. If we choose three most important words for every topic distribution, then the words in topic A might be: "Biden", "senate", and "government" while the most important words in topic B are: "software", "cloud", and "computer". The common words in topic C could be "movie", "music", "art". From there, we can see that topic A is clustered around politics keywords, and topic B talks about tech, while topic C is clustered around entertainment. Latent Dirichlet Allocation [26] is an unsupervised generative and probabilistic model for topic modeling which is widely used to extract topics from a corpus of textual documents. LDA represents each document

of the corpus as a distribution over latent topics, and each topic, in turn, is represented as a distribution over words. In LDA terminology, a document is represented as a sequence of N words denoted by w = (w_1,... w_n), where w_n is the nth word in the sequence, and a corpus is a collection of "M" documents denoted by D = W1...Wm. The generative process for each document "w" in the corpus "D" is given as follows [26]:

1. Choose $N \sim \text{Poisson}(\xi)$
2. Choose $\theta \sim \text{Dir}(\alpha)$
3. For each word w_n in the document sequence:

    3.1. Choose a topic $z_n \sim \text{Multinomial}(\theta)$
    3.2. Choose a word w_n from $p(w_n \mid z_n, \beta)$, a multinomial probability conditioned on the topic z_n

A k-dimensional Dirichlet random variable θ can take values in the $(k-1)$ simplex (a k-vector θ lies in the $(k-1)$ simplex if $\theta_i \geq 0, \sum_{i=1}^{k}\theta_i = 1$), and has the following probability density on this simplex:

$$p(\theta \mid \alpha) = \frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\theta_1^{\alpha_1-1}\cdots\theta_k^{\alpha_k-1} \tag{1}$$

where the parameter α is a k-vector with components $\alpha_i > 0$, and where $\Gamma(x)$ is the Gamma function. Given α and β parameters, the joint distribution of a topic mixture θ, a set of N topics $\mathbf{z}$, and a set of N words $\mathbf{w}$ is given by:

$$p(\theta, \mathbf{z}, \mathbf{w} \mid \alpha, \beta) = p(\theta \mid \alpha)\prod_{n=1}^{N}p(z_n \mid \theta)p(w_n \mid z_n, \beta) \tag{2}$$

where $p(z_n \mid \theta)$ is simply θ_i for the unique i such that $z_n^i = 1$. Integrating over θ and summing over z, we obtain the marginal distribution of a document:

$$p(\mathbf{w} \mid \alpha, \beta) = \int p(\theta \mid \alpha)\left(\prod_{n=1}^{N}\sum_{z_n}p(z_n \mid \theta)p(w_n \mid z_n, \beta)\right)d\theta \tag{3}$$

Finally, taking the product of the marginal probabilities of single documents, the probability of a corpus is computed by:

$$p(D \mid \alpha, \beta) = \prod_{d=1}^{M}\int p(\theta_d \mid \alpha)\left(\prod_{n=1}^{N_d}\sum_{d_dn}p(z_{dn} \mid \theta_d)p(w_{dn} \mid z_{dn}, \beta)\right)d\theta_d \tag{4}$$

6 CaMeL Analyzer for Sentiment Classification

Sentiment classification is the task of classifying texts into three main polarities: positive, neutral, and negative. In the process of sentiment classification of texts, three primary approaches are used: machine learning, lexicon-based approach, and hybrid approach [27]. In the machine learning approach, many algorithms are applied using linguistic features to detect the sentiment of the text. In this work, we used a machine learning-based pre-trained model to classify topic-categorized comments. However, Arabic sentiment classification remains a very challenging task due to many reasons. Some of them are related to the special nature of the language (mentioned in Section 4). Many researchers have put their efforts to develop resources and tools to solve the state-of-the-art opinion mining problems in the Arabic language. CaMeL Tools [28] is one of the essential projects aiming to provide an open-source toolkit of NLP tasks for the Arabic language such as named entity recognition, preprocessing tools, sentiment analysis, dialect identification, etc. CaMeL sentiment analyzer was pre-trained on large and various datasets. The classifier labels a given text into three polarities: positive, neutral, and negative. The analyzer using HuggingFace's Transformers [29] and fine-tuned on the top of AraBERT [30] and multilingual BERT [31] models to classify sentiments. The models were trained and tested on various datasets such as the dialectical dataset of Arabic Speech-Act and the sentiment corpus of tweets (ArSAS).

7 Topic Modeling-based Sentiment Analysis

In this section, we introduce our hybrid approach for topic-based sentiment analysis for Arabic text. Baseline methods in Arabic sentiment analysis (ASA) are used to simply compute opinion polarity for a given text in a document ignoring the topic (or aspect) level that lies beneath. However, we often find various topics are expressed in a subjective text and the text may also hold different polarities regarding these topics. Our approach intends to retrieve people's feedback towards various topics related to the COVID-19 pandemic.

The system aims, in the first phase, to extract topics within the corpus using a topic modeling method, namely LDA, and then reconstructs the words of each topic based on their probabilities in each topic. The topic words are then mapped to corresponding aspects using a logical operator. To enhance the formulated topics with other similar words in the collection, we trained

a word embedding model on the dataset using word2vec [32] to compute word similarities. The final formulated aspects (or topics) from the corpus are shown in Table 2, and the topic distribution proceeding from the topic extraction task is shown in Figure 1.

Table 2   Topic words for each extracted topic

Topic	Topic Words
Covid19 & Economy	أرباب-شركات-اغلاق-عمل-مصانع-اقتصاد-معامل- صناعة-شغل-متاجر-محلات
Covid19 & E-Learning	تلميذ-وزارة-تربية-منصة-مدرسة- امتحان-استاذ-صحي-حجر-عن بعد-عمومي-تعليم-دراسة
Covid19 & Fear	جوع-ملل-خطر-موت-بلاء-قلق-خوف-عدوى-سلامة-شفاء-الله-سقم-انتحار-مصيبة-تشاؤم
Covid19 & Government	حظر-العثماني-مواطن-حكومة-دولة-التزام-طوارئ- داخلية-صحي-حجر-رفع-تمديد
Covid19 & Health	كمامة-إصابات -تحاليل-نفسية-صحة -حالات-وقاية-تعقيم
Covid19 & support	مساعدات-ضمان-بطالة-ازمة-دعم-صندوق-راميد-تعويضات
Covid & Vaccine	لجنة-علمية-جواز-دلتا-لقاح-كوفيد-اوميكرون-سينوفارم-استرازينيكا-فايزر-فيروس-تلقيح-متحور-الغاء-جرعة
Covid19 & Misinformation	تأثير-ابادة-صينية-نظرية-ماسونية-مؤامرة-امريكية-اوهام-كذب-مسرحية-خطط

The extraction task resulted to distinguish 8 common topics related to COVID-19 pandemic:

- Economy: related to all domains of the economy affected by COVID-19 pandemic in Morocco such as factories, companies, tourism, agriculture, trade, work, industry, etc.
- Vaccine: relates to all aspects of COVID-19 vaccine and vaccination campaign such as vaccine, Sinopharm, AstraZeneca, Pfizer, variants of coronavirus (Delta, Omicron), dose, etc.
- E-learning: has a relation to online learning during quarantine periods because of COVID-19 pandemic. The topic is related to aspects like student, teacher, online platform, ministry of education, exams, etc.
- Support plan: relates to all aspects of support plan launched by the government to support employees who lost their jobs because of the economy shutdown like, joblessness, aids, crisis, indemnities, RAMID (an aid system launched by health ministry) etc.
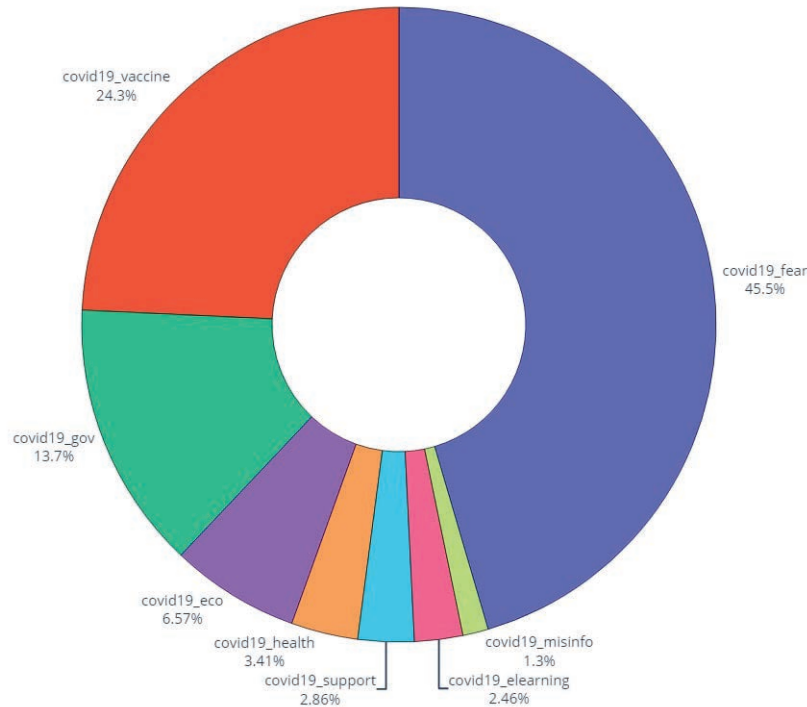
Figure 1   Extracted topics distribution.

- Health: relates to all aspects of health regarding COVID-19 pandemic such as masks, tests, social distancing, mental health, cases, prevention etc.
- Fear and anxiety: this topic relates to all aspects of fear of being infected, anxiety, suppress, psychological disorders during quarantine, dystopian feelings, etc.

In the second phase, we perform a full scan through the dataset and search for comments expressing the extracted topic. The Algorithm 1 receives the extracted topic as input and retrieves a data frame of comments expressing the input topic. This task is repeated for other topics. As a result, topic categorized data frames are constructed for sentiment classification task. The Algorithm 2 is applied on the topic-categorized generated data frames and then a topic-based polarity is computed using CaMeL pre-trained model [28]

to classify sentiments to negative, neutral, or positive with respect to these topics.

Algorithm 1 Topic extraction task

Data: comments
Result: topic_dataframes
    Preprocessed_comments ← clean_diacritics()&stopwords_romoval()&prprocess();
    Extracted_topics ← LDA(Preprocessed_comments);
    Enhanced_topics ← word2vecSimilarity(Extracted_topics);
    foreach topic **do**
        scan comment;
        if comment 'expresses' topic **then**
            topic_dataframe ← add(comment);
        else
            continue;
        end if
    end foreach

Algorithm 2 Topic-based sentiment analysis task

Data: topic_dataframes
Result: topic-based sentiment classification
    foreach topic_dataframe in topic_dataframes **do**
        foreach comment in topic_dataframe **do**
            sentiment_score ← CaMeLSentimentAnalyzer(comment);
            if sentiment_score > 0 **then**
                sentiment ← positive;
            else if sentiment_score < 0 **then**
                sentiment ← negative;
            else
                sentiment ← neutral;
            end if
        end foreach
    end foreach

The overall process of our work, as illustrated in Figure 2, starts with collecting the comments of articles from Hespress and then saved as data frame csv file. Afterwards, cleaning and preprocessing tasks are performed on the raw comments. Afterwards, topics are extracted within the collection of documents (text comments) using LDA. For each topic, we map the corresponding comments using a regular expression searching method. As a result, the comments are assigned to related topics. Finally, we perform
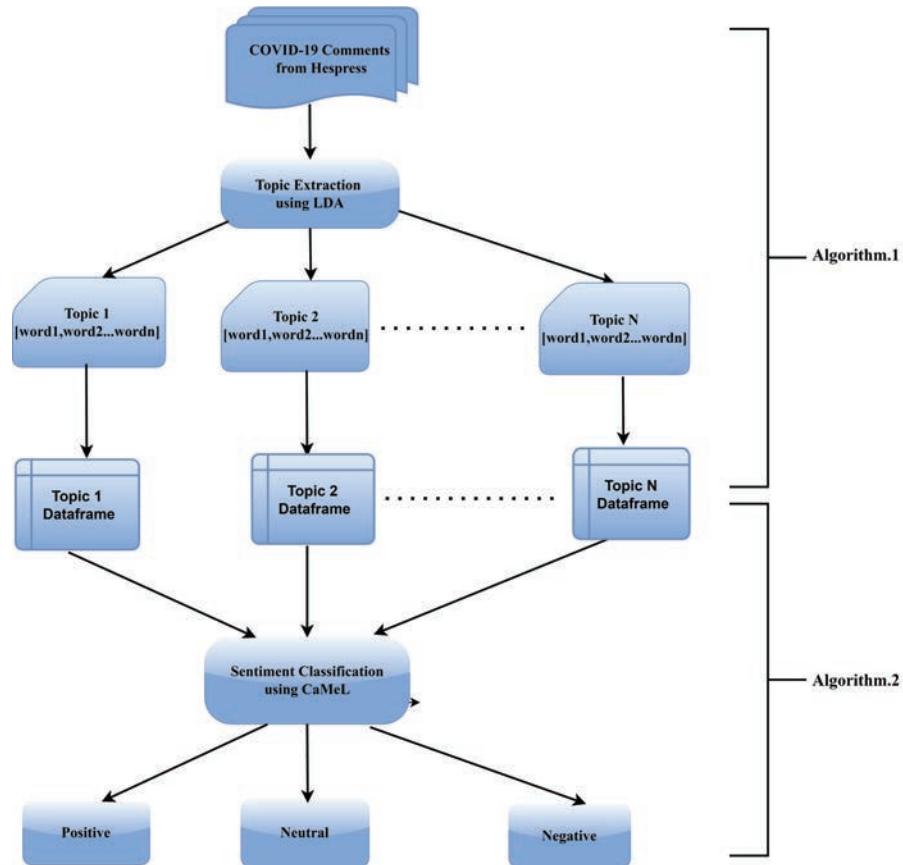
Figure 2   Topic-based sentiment analysis workflow.

sentiment analysis on a topic-based level. The final results of the overall process are shown and discussed in the next section.

8 Results and Discussion

In this section, we present and discuss the outcomes of our study. Topic-based sentiments regarding all extracted topics are shown in Table 3, and the frequency of each polarity for each topic, as well as the rate of the polarity (the polarity rate is computed by dividing the sentiment frequency by the sum of all frequencies of other sentiments for a specific topic).

Table 3   Topic-based sentiment analysis results

Topic	Sentiment	Frequency	Rate (%)
Covid19 & Economy	Positive	182	7.34
	Neutral	522	15.1
	Negative	2290	77.6
Covid19 & Government	Positive	571	11.40
	Neutral	774	13.2
	Negative	4886	75.5
Covid19 & E-Learning	Positive	115	7.23
	Neutral	274	25.3
	Negative	743	67.5
Covid19 & Fear	Positive	656	13.3
	Neutral	303	32.2
	Negative	1109	54.5
Covid19 & Health	Positive	122	5.48
	Neutral	382	37.0
	Negative	1530	57.5
Covid19 & Support	Positive	82	3.77
	Neutral	380	17
	Negative	980	79.2
Covid19 & Vaccine	Positive	42	3.8
	Neutral	262	23.7
	Negative	800	72.5
Covid19 & Misinformation	Positive	92	5.48
	Neutral	132	17.8
	Negative	560	76.7

The obtained results support our first assumption that the COVID-19 pandemic has affected dramatically all aspects of normal life. All those expressed sentiments are highly negative regarding all topics. We notice that the most of respondents feel deficient on how this pandemic affected the economy. The majority of them also expressed fear, worry, and concerns about losing their employments, and those who had lost their jobs were dissatisfied with the government's support plan during the quarantine months. The reported findings in Table 3 and Figure 3 revealed a high negative rate (around 79%) of expressed sentiments regarding COVID-19 & support topics. We also highlight that the highest negative rate of sentiments expressed fear and anxiety towards the new coronavirus, which is understandable, given that all the comments were gathered in the early months of the pandemic where the virus was still considered an unknown threat to the living humans.
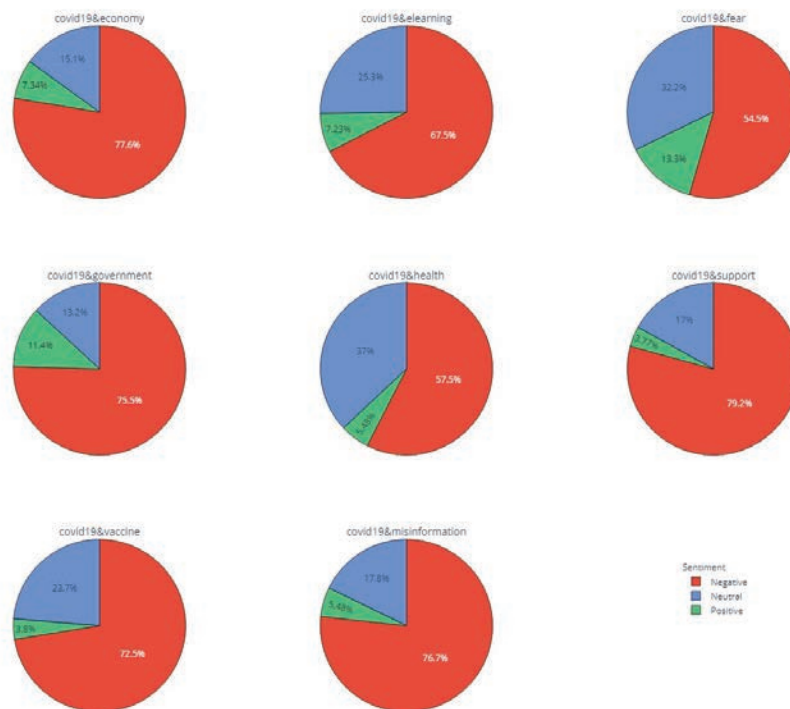
Figure 3 Topic-based sentiment analysis pie chart results.

We notice that the majority (72%) of comments related to vaccine and vaccination are negative which is understandable if knowing that many people still hold feelings of concern and worry towards COVID-19 vaccine. We also observe that a high rate (76%) of comments related to COVID-19 & misinformation is highly negative. The way of fake news and misleading information about the COVID-19 pandemic and the vaccine are spreading in social media and the web has made people to develop feelings of fear and mistrust on taking the vaccine. As a result, people started to form anti-vaccination groups to protest against the vaccination campaigns and the use of a mandatory vaccine pass to access governmental buildings, restaurants, theatres, work places, etc.

The major purpose to conduct this research paper was to analyze topic-based opinions which are expressed by Moroccan citizens about COVID-19 impact on several domains like labor market, education, economy, mental

health, etc. The outcomes were designed to provide an answer to the question of how this major crisis impacted people's everyday lives. Assuming that this study needs to be carried out with more interest in terms of choosing the right methods and tools and to be applied on a large-scale data,then the results would be much coherent and insightful to cover other domains that are not mentioned in the current study.

COVID-19 pandemic has affected calamitously education and made it inaccessible for the majority of students, especially those who are living in the countryside and suffering from lack of technological means (internet infrastructure, laptops, tablets, etc.) that could help them to pursue their classes online.

9 Conclusion and Future Work

This paper proposed an approach for topic-based sentiment analysis in the field of Arabic sentiment analysis. The system combines a topic modeling method (LDA) with sentiment analysis pre-trained analyzer (CaMeL) applied on COVID-19 collected comments to perform opinion mining on a topic based-level. The obtained results showed that all expressed sentiments regarding all extracted topics were highly negative. We propose as future works, to expand this study to cover other neighboring countries in north Africa to mine the impact of COVID-19. We also suggest to conduct another study on finding solutions for the challenges of e-learning in Morocco when face-to-face learning is avoidable like in the situation of COVID-19. One of these challenges is the protection of users' privacy, and we plan on using homomorphic [33] encryption as a solution for its promising previous results [34, 35].

Acknowledgments

References

[1] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, A Novel Coronavirus from Patients with Pneumonia in China, 2019, N. Engl. J. Med. (2020). https://doi.org/10.1056/nejmoa2001017.

[2] R. Karia, I. Gupta, H. Khandait, A. Yadav, A. Yadav, COVID-19 and its Modes of Transmission, SN Compr. Clin. Med. (2020). https://doi.org/10.1007/s42399-020-00498-4.

[3] Brooks, Sam, Webster, Rebecca, Smith, Louise, Woodland, Lisa, Wessely, Simon, Greenberg, Neil, Rubin, Gideon. (2020). The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The Lancet. https://doi.org/10.1016/S0140-6736(20)30460-8

[4] Passavanti, Marco, Argentieri, Alessandro, Barbieri, Diego, Lou, Baowen, Wijayaratna, Kasun, Foroutan Mirhosseini, Ali, Fusong, Wang, Naseri, Sahra, Qamhia, Issam, Tangerås, Marius, Pelliciari, Matteo, Ho, Chun-Hsing. (2021). The Psychological Impact of COVID-19 and Restrictive Measures in the World. Journal of Affective Disorders. https://doi.org/10.1016/j.jad.2021.01.020

[5] El-Ansari, A., Beni-Hssane, A., Saadi, M. (2020, March). An improved modeling method for profile-based personalized search. In Proceedings of the 3rd international conference on networking, information systems & security (pp. 1–6).

[6] El-Ansari, A., Beni-Hssane, A., Saadi, M. (2020). An ontology-based profiling method for accurate web personalization systems. Journal of Theoretical and Applied Information Technology, 98(14), 2817–2827.

[7] Birjali, Marouane, Kasri, Mohammed, Beni hssane, Abderrahim. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-Based Systems. https://doi.org/10.1016/j.knosys.2021.107134

[8] El-Ansari, A., Beni-Hssane, A., Saadi, M., El Fissaoui, M. (2021). PAPIR: privacy-aware personalized information retrieval. Journal of Ambient Intelligence and Humanized Computing, 1–17.

[9] El-Ansari, A., Beni-Hssane, A., Saadi, M. (2017). A multiple ontologies based system for answering natural language questions. In Europe and MENA cooperation advances in information and communication technologies (pp. 177–186). Springer, Cham.

[10] El-Ansari, A., El Fissaoui, M., Beni-Hssane, A. (2021, April). Personalized question-answering over linked data. In Proceedings of the

4th International Conference on Networking, Information Systems & Security (pp. 1–7).

[11] P. Melville, W. Gryc, R. D. Lawrence, "Sentiment analysis of blogs by combining lexical knowledge with text classification", In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1275–1284, ACM, 2009. https://doi.org/10.1145/1557019.1557156

[12] Abdul-Mageed, Muhammad, Sandra Kübler, Mona Diab. "Samar: A system for subjectivity and sentiment analysis of arabic social media." In Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, pp. 19–28. Association for Computational Linguistics, 2012. https://doi.org/10.1016/j.csl.2013.03.001

[13] X. Hu, J. Tang, H. Gao, H. Liu, "Unsupervised Sentiment Analysis with Emotional Signals", Rio de Janeiro, Brazil ACM 978-1-4503-2035-1/13/05, 2013. https://doi.org/10.1145/2488388.2488442

[14] Dasgupta, Vincent Ng, "Topic-wise, sentiment-wise, or otherwise?: Identifying the hidden dimension for unsupervised text classification", In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2, pp. 580–589. Association for Computational Linguistics, 2009. https://dl.acm.org/doi/10.5555/1699571.1699589

[15] Taoufiq, Zarra, Chiheb, Raddouane, Moumen, Rajae, Faizi, Rdouan, El Afia, Abdellatif. (2017). Topic and sentiment model applied to the colloquial Arabic: a case study of Maghrebi Arabic. https://doi.org/10.1145/3128128.3128155

[16] N. Shelke, S. Deshpande, V. Thakare, Domain independent approach for aspect oriented sentiment analysis for product reviews, in: Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications, pp. 651–659, Springer, Bhubaneswar, India, 2017. https://doi.org/10.1007/978-981-10-3156-4_69

[17] J. J. Li, H. Yang and H. Tang, Feature mining and sentiment orientation analysis on product, in: Management Information and Optoelectronic Engineering: Proceedings of the 2016 International Conference on Management, Information and Communication (ICMIC2016) and the 2016 International Conference on Optics and Electronics Engineering (ICOEE2016), p. 79, World Scientific, Guilin, China, 2017. https://doi.org/10.1142/9789813202689_0011

[18] M. S. Akhtar, D. Gupta, A. Ekbal, P. Bhattacharyya, Feature selection and ensemble construction: a two-step method for aspect-based sentiment analysis, Knowledge- Based Systems. https://doi.org/10.1016/j.knosys.2017.03.020

[19] R. Piryani, V. Gupta, V. K. Singh and U. Ghose, A linguistic rule-based approach for aspect-level sentiment analysis of movie reviews, in: Advances in Computer and Computational Sciences, pp. 201–209, Springer, Singapore, 2017. https://doi.org/10.1007/978-981-10-3770-2_19

[20] V S, Anoop, S., Asharaf. (2018). Aspect-Oriented Sentiment Analysis: A Topic Modeling-Powered Approach. Journal of Intelligent Systems. 29. 10.1515/jisys-2018-0299. https://doi.org/10.1515/jisys-2018-0299

[21] Hankar M., Birjali M., El-Ansari A., Beni-Hssane A. (2022) Arabic Topic Modeling-Based Sentiment Analysis on COVID-19 Feedback Comments. In: Maleh Y., Alazab M., Gherabi N., Tawalbeh L., Abd El-Latif A.A. (eds) Advances in Information, Communication and Cybersecurity. ICI2C 2021. Lecture Notes in Networks and Systems, vol. 357. Springer, Cham. https://doi.org/10.1007/978-3-030-91738-8_9

[22] Elena Rudkowsky, Martin Haselmayer, Matthias Wastian, Marcelo Jenny, Štefan Emrich & Michael Sedlmair (2018): More than Bags of Words: Sentiment Analysis with WordEmbeddings, Communication Methods and Measures. https://doi.org/10.1080/19312458.2018.1455817

[23] Erritali, M., Beni-Hssane, A., Birjali, M., Madani, Y. (2016). An approach of semantic similarity measure between documents based on big data. International Journal of Electrical and Computer Engineering, 6(5), 2454.

[24] Hegazi, Mohamed Osman, Al-Dossari, Yasser, Al-Yahy, Abdullah, Al-Sumari, Abdulaziz, Hilal, Anwer. (2021). Preprocessing Arabic text on social media. Heliyon. https://doi.org/10.1016/j.heliyon.2021.e06191

[25] Ghadah Alwakid, Taha Osman, Thomas Hughes-Roberts.(2017). Challenges in Sentiment Analysis for Arabic Social Networks. Procedia Computer Science. https://doi.org/10.1016/j.procs.2017.10.097.

[26] Blei, David, Ng, Andrew, Jordan, Michael. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research. https://jmlr.csail.mit.edu/papers/v3/blei03a.html

[27] Maynard, Diana, Funk, Adam. (2011). Automatic Detection of Political Opinions in Tweets. CEUR Workshop Proceedings. https://doi.org/10.1007/978-3-642-25953-1_8

[28] Obeid, Ossama, Zalmout, Nasser, Khalifa, Salam, Taji, Dima, Oudah, Mai, Alhafni, Bashar, Eryani, Fadhl, Erdmann, Alexander, Habash, Nizar. (2020). CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing. https://aclanthology.org/2020.lrec-1.868.pdf

[29] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Brew, J. (2019). Huggingface's transformers: State-of-the-art natural language processing. https://arxiv.org/abs/1910.03771

[30] Antoun, Wissam, Baly, Fady, Hajj, Hazem. (2020). AraBERT: Transformer-based Model for Arabic Language Understanding. https://arxiv.org/abs/2003.00104

[31] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. https://arxiv.org/abs/1810.04805

[32] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space, 2013, pp. 1–12. https://arxiv.org/abs/1301.3781

[33] El Makkaoui, K., Beni-Hssane, A., Ezzati, A. (2019). Speedy Cloud-RSA homomorphic scheme for preserving data confidentiality in cloud computing. Journal of Ambient Intelligence and Humanized Computing, 10(12), 4629–4640.

[34] El Makkaoui, K., Ezzati, A., Beni-Hssane, A., Motamed, C. (2016, May). Cloud security and privacy model for providing secure cloud services. In 2016 2nd international conference on cloud computing technologies and applications (CloudTech) (pp. 81–86). IEEE.

[35] El Makkaoui, K., Beni-Hssane, A., Ezzati, A. (2016, October). Cloud-ElGamal: An efficient homomorphic encryption scheme. In 2016 International Conference on Wireless Networks and Mobile Communications (WINCOM) (pp. 63–66). IEEE.

Biographies

Mustapha Hankar is a PhD student researcher at the department of computer science, Faculty of Science, University of Chouaib Doukkali, El Jadida, Morocco since 2019. His research areas include NLP and its applications, machine learning and deep learning. He contributed in the proceedings of the international conferences.

Marouane Birjali received his Ph.D in computer science from the Faculty of Sciences, Chouaïb Doukkali University, El Jadida since 2019. Currently, he is a Researcher in the same faculty and working as an IT engineer. His research interests include Big Data, IA and Sentiment Analysis is a researcher in the field of NLP and Big Data and AI. He contributed to many national and international conferences, and published many papers in several international journals in his research area.

Anas El-Ansari received his PhD. degree in computer science from the Faculty of Sciences, Chouaïb Doukkali University, El Jadida. Currently, he is a Researcher and a Professor in the Polydisciplinary Faculty of Nador, Mohamed First University, Morocco. His research interests include Recommender systems, Cryptography, Privacy, Sentiment Analysis and Semantic Web. He contributed and published papers in many national, international conferences and journals.

Abderrahim Beni-Hssane is a researcher and professor of Computer Science in the Faculty of sciences at the university of Chouaib Doukkali, El Jadida, Morocco. He received his Ph.D degree in computer science from Mohamed V University, Rabat, Morocco, in 1997. His research interests include performance evaluation in wireless networks, cryptography, cloud computing, big data, machine learning, and NLP. He contributed and published papers in many national, international conferences and journals.