

Received June 25, 2021, accepted July 10, 2021, date of publication July 15, 2021, date of current version July 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3097492

Sentiment Analysis of Political Tweets From the 2019 Spanish Elections

MARGARITA RODRÍGUEZ-IBÁÑEZ¹, FRANCISCO-JAVIER GIMENO-BLANES²,
PEDRO MANUEL CUENCA-JIMÉNEZ³, CRISTINA SOGUERO-RUIZ³,
AND JOSÉ LUIS ROJO-ÁLVAREZ^{3,4}, (Senior Member, IEEE)

¹Department of Business, Universidad Rey Juan Carlos, 28933 Madrid, Spain

²Department of Communication Engineering, Universidad Miguel Hernández, Elche, 03202 Alicante, Spain

³Department of Signal Theory and Communications, Telematics, and Computing Systems, Universidad Rey Juan Carlos, 28933 Madrid, Spain

⁴Center for Computational Simulation, Universidad Politécnica de Madrid, 28040 Madrid, Spain

Corresponding author: José Luis Rojo-Álvarez (joseluis.rojo@urjc.es)

This work was supported in part by the KLINILYCS under Grant TEC2016-75361-R, in part by the AAVis-BMR under Grant PID2019-107768RA-I00, in part by the BigTheory under Grant PID2019-106623RB-C41, in part by the meHeart RisBi under Grant PID2019-104356RB-C42, in part by the miHeart-DaBa through the Agencia Estatal de Investigación and co-funded by FEDER under Grant PID2019-104356RB-C43, in part by the Young Researchers Research and Development Project through the Community of Madrid and Rey Juan Carlos University under Project 2020-656, in part by the KERMES under Grant TEC2016-81900-REDT, and in part by the MAPAS through the Spanish Government under Grant TIN2017-90567-REDT.

ABSTRACT The use of sentiment analysis methods has increased in recent years across a wide range of disciplines. Despite the potential impact of the development of opinions during political elections, few studies have focused on the analysis of sentiment dynamics and their characterization from statistical and mathematical perspectives. In this paper, we apply a set of basic methods to analyze the statistical and temporal dynamics of sentiment analysis on political campaigns and assess their scope and limitations. To this end, we gathered thousands of Twitter messages mentioning political parties and their leaders posted several weeks before and after the 2019 Spanish presidential election. We then followed a twofold analysis strategy: (1) statistical characterization using indices derived from well-known temporal and information metrics and methods –including entropy, mutual information, and the Compounded Aggregated Positivity Index– allowing the estimation of changes in the density function of sentiment data; and (2) feature extraction from nonlinear intrinsic patterns in terms of manifold learning using autoencoders and stochastic embeddings. The results show that both the indices and the manifold features provide an informative characterization of the sentiment dynamics throughout the election period. We found measurable variations in sentiment behavior and polarity across the political parties and their leaders and observed different dynamics depending on the parties' positions on the political spectrum, their presence at the regional or national levels, and their nationalist or globalist aspirations.

INDEX TERMS Sentiment analysis, twitter, social networking sites, text analysis, lexicon, politics, collected tweets, election candidates, election results, manifold embedding, autoencoders.

I. INTRODUCTION

It is no secret that in an increasingly connected society, social media plays a growing role in the way people interact. Consequently, social networks have become a relevant representation of social sentiment, not only because of the opinions expressed and recorded on them but also because of their emotional impact on the readers. This is why a large number of recent studies on sentiment analysis have focused on social

media environments. In the case of Twitter, approximately two hundred papers are published annually that include the keywords *sentiment* and *twitter* in their titles [1]–[3]. One reason for this large amount of research activity in the area of social media is the cost-effectiveness of gathering large datasets for scrutinizing the information conveyed by its different platforms. Only a few years ago, the only way to incorporate such extensive datasets in a study was through effort-intensive tasks such as social polls or surveys. Now, the reverse is true, as new technologies can add novel, powerful ways to craft extensive datasets [4]. According to the

The associate editor coordinating the review of this manuscript and approving it for publication was Fabrizio Marozzo^{id}.

literature, these efforts are not restricted to a few niche topics that might be of temporary interest. On the contrary, the scope of works in this area encompasses a vast number of disciplines and activities from many unrelated fields, including commercial activities between customers and manufacturers, the analysis of stock prices, the satisfaction of people supported by a given nongovernmental organization, or the expectations of voters of a certain political party, to name a few [5]–[7].

The increasing information volume that can be easily retrieved from social media, its multimodal nature, and the complexity of modeling sociological and psychological attributes associated with sentiment have together led to the use of machine learning as one of the most effective tools for this type of analysis [8]–[10]. This approach is encouraged by the great success that has been achieved with machine learning techniques in many disciplines [11]–[13], including sentiment analysis, which is mostly supported today with the combination of machine learning and natural language processing [14]. Particular attention has been given to electoral research, such as prediction of the political sentiments of voters from short messages taking into account multiparty contexts in India [15], the detection of emerging political topics during German parliamentary elections comparing short messages with Google trends [16], or the evidence of negative sentiment in speeches from different actors from the US presidential election [17]. Some studies have obtained nontrivial information regarding election dynamics, such as the more persistent spreading of negative information than positive or neutral information about candidates, polarization in terms of sentiments spread by followers, and even the spread of misinformation by followers of the winner [18]. Surprisingly, few studies have aimed to scrutinize the temporal evolution of the dynamics of sentiment in electoral scenarios, probably because most of the existing analysis schemes use machine learning engines (such as logistic regression or long short-term memory networks) and features, which cannot be easily adapted to follow the temporal evolution of these dynamics.

Accordingly, in this study, we performed out a multivariate dynamic analysis of sentiment extracted from social media to characterize its properties and evolution in occasions of singular social interest, particularly during the period around the elections of national representatives. For this purpose, tweets related to the main political leaders and their parties were collected starting a month before and up to one week after the presidential elections in Spain. Building on models developed and validated in previous works [3], we then performed an in-depth analysis of the statistical properties of these data. We started by consolidating and validating various lexicons on which we based our sentiment extraction process, as this is fundamental for any subsequent analysis. Then, a twofold strategy was followed: first, we selected some indices mostly based on information theory, which *a priori* could allow us to retrieve informative knowledge about the dynamics and statistical evolution of the sentiment. In this setting, variables such as sample entropy, mutual information, and a

recently proposed positivity index [3] were scrutinized, and several experiments were performed to explore the dynamics of these variables throughout the election period. Second, we aimed to analyze in detail the information retrieved by intrinsic nonlinear mappings of input spaces to embedded spaces, as this represents an intrinsic and likely nonlinear feature extraction. In this direction, we examined two of the most widely used embedding schemes, namely, t-distributed stochastic neighbor embedding (tSNE) and the autoencoder architecture, to study the interpretability of the embeddings to obtain a more detailed and complete description perspective on the data.

Most current political parties use social media networks constantly (Twitter, Facebook, and Instagram, among others) to manage, get closer to and determine their voters' opinions. However, there is moderate public scientific evidence regarding the information extraction systems that different parties use or the content of the messages they extract from the various networks. Considering this gap and the available tools in this field, the main contributions of this paper are (a) the exploration of the information extraction capacity of artificial intelligence tools when applied to sentiment in social networks, with a focus on the analysis of temporal dynamics from simple summary indices; (b) the identification of the type of information that parties can find on Twitter and the way that it is reported; and (c) the exploratory study of the temporal dimension of messages posted on Twitter before, during and after a political event of great magnitude in Spain, such as the Spanish 2019 presidential elections.

The rest of this paper is organized as follows. Section II presents the relevant background on sentiment analysis and the method evaluated in the literature. Section III introduces the methods used and the statistical dynamics established in this paper. Section IV describes the experiments performed and the subsequent results obtained. Finally, Section V provides additional discussion and offers a set of conclusions.

II. BACKGROUND

Sentiment analysis has been shown to be a remarkable tool in almost every research area [19]. As such, a large number of studies have been published on the subject [20]. Sentiment analysis has been applied to social media in an attempt to capture political polarity and orientation [21]–[23]; some of these results have shown [21], [24]–[26] that social media content can reliably predict certain political events and the development of political campaigns.

Our research here is focused on *Twitter*, an extended microblogging platform whose users share short messages called *tweets* that can include links to other websites, videos or pictures [27]. Given that politics is a rich source of controversial views, it is a particularly active topic on this social network. The application of sentiment analysis techniques to the Twitter corpus is a popular research subject today, and many studies have attempted to use these techniques in the context of political sentiment analysis and extraction [28].

This section briefly reviews previous studies on the topic of sentiment analysis of political messaging. We first explore the landscape of political sentiment analysis techniques that rely on Twitter and then discuss the main methods used to extract sentiment information from written content.

A. RELATED WORKS

The political orientation of Italian users on Twitter was predicted using two machine learning techniques, decision trees and K-nearest neighbors, and a specifically developed naive Bayes multinomial text classification algorithm [26]. Additionally, trends in the Venezuelan parliamentary election of December 2015 were predicted [24] using a sentiment analysis technique based on unsupervised machine learning, together with Linguistic Inquiry and Word Count software to automatically analyze tweets for various linguistic traits. Similar techniques were applied by [29], whose authors inferred the results of the general elections of India in 2019 with a long short-term memory (LSTM) methodology, which was also compared with classical machine learning models.

However, many researchers have concluded that they cannot predict the outcome of political events [30]–[32]. Using natural language processing (NLP), the negative and positive polarities in tweets were found to be insufficient indicators for determining support for any of the candidates in the US presidential election of 2016.

This topic remains open to discussion, given that in the same context (the US presidential election of 2016), other authors [33], using a lexicon-based approach, determined the polarity and bias measures of the collected tweets. They found that public opinion about the candidates did have an impact on the potential elected leader for the country. Both [33] and [32] used word cloud plots to represent the words that appeared most frequently in tweets, and both of them collected tweets and performed political sentiment analysis using a four-step approach: retrieval, preprocessing, analysis, and visualization of the results. The former study noted that a key requisite for this analysis was the possession of a good-quality sentiment dictionary as the research foundation. We also encountered this issue in our research, as most dictionaries are originally in English or other languages; the difference between a dictionary translated into Spanish and a dictionary created in Spanish, for example, is notable.

In another study, the sentiment polarity of political tweets was analyzed during the Spanish presidential elections in 2015 and 2016 [34]. There, natural language processing and text mining techniques were implemented based on the LinguaKit dictionary, and a model assigned tweets to one of three categories (positive, negative or neutral). Other authors have used emoticons to determine the polarity of tweets and assign sentiment scores to online texts [35].

In a different line of research, the impact of malicious accounts on political tweet sentiment was explored [36], [37]. Some of these approaches have proposed semisupervised methods (using word embeddings) to classify Twitter

hashtags and detect groups of interest and potential thread hijackers commenting on political events. Others have built datasets using convolutional neural networks trained on the *sentiment140 dataset* [36]. In another example of the use of word embeddings [38], a neural network was trained to learn word cooccurrences and generate word vectors from a corpus of four million political tweets extracted during the EU referendum of 23 June 2016. With a similar scope, [39] attempted to successfully estimate the polarization of social media users in the 2018 Italian general election and the 2016 US presidential election using a new methodology called Iterative Opinion Mining using Neural Networks (IOM-NN). Additionally, [40] proposed a cloud-based algorithm to perform similar analyses.

Other important contributions to the analysis of political sentiment have followed both different and related directions. In [27], a machine learning algorithm was created to classify the polarity of political tweets and to build a Corpus of Spanish Tweets (COST). In [41], the authors concluded that the results of many political sentiment studies depend on the quality of the dictionaries used. To investigate this aim, the authors compiled new lexicons using data from the Summer 2015 referendum in Greece, on which advanced techniques such as sarcasm correction were employed. In [42], the researchers created a specific political sentiment index to assess the views users posted on Twitter during the 2015 UK election. Finally, in [43], a Tweets Sentiment Analysis Model was proposed that could capture societal interest and people's opinions toward a specific event; the study demonstrated that building a lexicon-based sentiment analysis intelligent system could be very beneficial.

B. SENTIMENT ANALYSIS APPROACHES

Sentiment analysis, as defined in [44], is the action of explaining the meaning of emotions such as positive, negative, or neutral through various text-mining methods and materials. Sentiment analysis approaches can be divided into three different categories, namely, machine learning techniques, lexicon-based models, and hybrid approaches.

Machine learning techniques are used to create models that predict sentiment from text using classification algorithms [3]. Approaches based on machine learning can use supervised or unsupervised learning techniques to build a model from training data. In supervised learning, a (possibly nonlinear) function is learned by mapping input and output pairs through the use of labeled data to drive the learning process. Unsupervised learning, on the other hand, draws inferences from unlabeled datasets [45]. Some machine learning techniques commonly used for sentiment analysis include the following:

- 1) NLP methods, which generally aim to help people communicate with computers through the use of natural language. Some of the techniques used in NLP can be applied to extract, analyze, and categorize information from textual data.

- 2) Case-based reasoning (CBR) techniques are designed to solve a new problem by remembering a previous, similar situation [46]. Sentiment analysis can arguably be understood as a knowledge-based classification problem [47].
- 3) Artificial neural networks, or simply neural networks, learn a function to understand and translate input data into the desired output [45], [48].
- 4) Support vector machines (SVMs) are supervised learning models used for classification and regression. In these techniques, a hyperplane is discovered that can be used to categorize points from the dataset in a high-dimensional space, thus leading to determination of the discriminant function that best separates the groups of data points [49].

Lexicon-based approaches, also known as symbolic approaches [32], use a sentiment dictionary to determine polarity, and sentiment scores are assigned to words to reflect the positive, negative, or neutral attitude of the speaker [50]. There are three main schemes among lexicon-based approaches [51], namely, manual schemes, dictionary-based schemes, and corpus-based schemes. Finally, hybrid approaches are combinations of machine learning and lexicon-based techniques.

III. MATERIAL AND METHODS

Several statistical models can be used to measure the accuracy of the multidimensional information that is presented and expressed in Twitter. In this section, we cover a wide range of such models, from more conventional approaches through temporal analysis and information dynamics, and then we present new state-of-the-art techniques based on autoencoders to illustrate the embedded manifolds underlying behavioral reality. Finally, the last subsection incorporates a description of the dataset used to perform this study.

A. TEMPORAL AND INFORMATION DYNAMICS

Tweets could ultimately be considered time-variant random processes, in which case they can be modeled as simplified stochastic processes for subsequent analysis of the multidimensional time evolution of the sentiment of certain population samples regarding a subject or topic. This approach allows the application of the corresponding and well-known theory of temporal series. In this work, we propose the use of window aggregation in the time domain to account for time-evenly distributed samples, and by doing so, we can define an appropriate environment for analyzing sentiment as a discrete-time process. The statistical analysis developed here, consistent with that in previous works [3], incorporates the detailed study of the variables intensity, feeling, their mean and standard deviation, as well as the calculation of entropy and mutual information from these variables. Some of them are summarized below for reader's convenience. In particular, the entropy (H) measures the weighted average value of the

amount of information ($H(j)$) and is calculated as follows:

$$H(j) = - \sum_{\rho} P(f, \rho) \log_2(P(f, \rho)), \quad (1)$$

where ρ stands for all unique possible sentiment scores, f is the entity, and P is the discrete probability density function of the score ρ for entity f . Therefore, $H(f)$ represents a one-dimensional array of the differential entropy evaluated in the time slot under analysis for every considered entity. Mutual information, meanwhile, measures the cross-dependence of two paired random variables.

Finally, in an attempt to make the evoked sentiment description independent of the number of issued elements, a new index was defined [3] in which, although the statistical reality of the parameter may vary in accordance with the volume of projected information, a one-to-one benchmark is established among users or topics, regardless of the difference in terms of volume of published messages. The Compounded Aggregated Positivity Index (CAPI) is defined as ten times the logarithm of relation among the compounded aggregated positive and negative sentiments and is mathematically computed as follows:

$$CAPI_f = 10 \cdot \log \left(\frac{\sum_{\rho>0} (\rho \cdot p_f[\rho])}{\sum_{\rho<0} (\rho \cdot p_f[\rho])} \right) (dB), \quad (2)$$

where $CAPI_f$ is the CAPI for user f (in dB), ρ represents all possible score values, and $p(\rho)$ is the value of the histogram of the score of the sentiment for ρ , in other words, the number of messages from user f where the sentiment score was evaluated as ρ .

B. MANIFOLD LEARNING AND AUTOENCODERS

Researchers must often contend with an interesting dilemma when dealing with high-dimensionality data. On the one hand, the attempt to represent the high-dimensional reality makes it almost impossible to begin to unravel the complexity and dynamism of the information, as it cannot be fully projected onto the proper dimensions for visual inspection. On the other hand, the reduction of dimensionality to two or three dimensions, which would allow visual inspection, can lead to a loss of information, preventing the ability to accurately describe and interpret the data. To address this dilemma, many supervised and unsupervised linear dimensionality reduction approaches have been proposed in the literature, such as principal component analysis (PCA) or linear discriminant analysis (LDA) [52]. However, although these frameworks could offer valid analyses in many environments and for many datasets, they usually are unable to capture the nonlinear patterns often present in natural data.

Manifold learning methods are considered a subgroup of nonlinear dimensionality reduction procedures [53], and as with any other variable reduction techniques, these methods are used under the hypothesis that high-dimensional data may actually be projected into a low-dimensional manifold, effectively representing the essential information of the original dataset. Therefore, reparametrizing the data according to

these manifolds can be useful for generating low-dimensional embeddings. Many approaches have been proposed in the literature in this regard, including isometric feature mapping (ISOMAP), local linear embedding (LLE), and the tSNE algorithm [53]. In essence, all these algorithms focus on extracting low-dimensional manifolds that can be designed to effectively describe high-dimensional data. Among them, in this paper, we evaluate tSNE, as well as some autoencoders. Briefly, tSNE is a nonlinear method widely used to visualize high-dimensional data in a low-dimensional space (two or three dimensions) [54]. From an algorithmic perspective, a probability distribution of high-dimensional samples is constructed so that similar samples have a high probability of being selected, while different samples have a much lower probability of being chosen. The tSNE algorithm defines similar distributions, from a stochastic perspective, for samples in the lower-dimensional space, and the Kullback–Leibler divergence between the two distributions is minimized with respect to the locations of the samples in the embedding. Although these techniques are often effective in describing the space and creating different embeddings, the main drawback lies in the lack of extension of the model to samples other than those previously introduced in the learning phase.

Another and even more useful alternative is the previously introduced manifold-based autoencoder set of methods [55]. These methods are also considered unsupervised deep learning generators of embeddings and are frequently used to learn latent representations from the data. Generally, an autoencoder is a neural network trained to set a target to be equal to the input. It typically consists of two main components: the encoder, used to map the input into a reduced space called the *latent space*; and the decoder, which maps the reduced space onto the reconstruction of the original input. To train an autoencoder, a cost function needs to be optimized by computing the difference between the training input vector x_i and its estimation \hat{x}_i . In other words, we can depict the autoencoder as a two-stage transformation, with the first stage involving an initial compression of variables and the second involving expansion toward full recovery of the original signal data. Although this process is clearly destructive, as full recovery of the original signal will not be possible due to the intermediate compression of the variables, by minimizing the difference between the original and final space, models defined in this way will eventually tend to preserve the essence of the data. Compared with those from tSNE, the models obtained from autoencoders can be better generalized over new samples and therefore extended over a different or incremental sample space to perform complete machine learning validation and testing processes.

The typical form of the encoder is an affine mapping represented by:

$$h_i = f(x) = \phi(W_e x_i + b_e), \quad (3)$$

where $h_i \in \mathbb{R}^d$ is the mapped vector in the latent space that corresponds to input $x_i \in \mathbb{R}^D$, $f(x)$ is the nonlinear transformation from the input to the latent space, $\phi(\cdot)$ is the

nonlinear activation, W_e is the weight matrix, and b_e is the bias. The corresponding decoder is also an affine mapping, given as:

$$z_i = g(h_i) = \varphi(W' h_i + b'), \quad (4)$$

where $g(\cdot)$ is the nonlinear transformation from the input space to the latent space, $\varphi(\cdot)$ denotes the nonlinear activation, W' is the weight vector, b' is the bias vector, and z_i is the estimated output. We want to estimate $f(\cdot)$ and $g(\cdot)$ from a set of samples and therefore the estimated weights and biases W , b , W' , b' such that $z_i = \hat{x}_i$.

C. POLITICAL SENTIMENT DATASET

To perform these tasks, and given the lack of open datasets available for the scope of this work, we developed a dataset for subsequent preprocessing and analysis. To create the necessary dataset, a previous successfully tested Sentiment Analysis Tool was used [3]. Although detailed information about the tool can be found in the referenced paper, for reader convenience, we state here that it works alongside the Twitter API called *twitty* to gather text data from the social network. The sentiment analysis performed by the tool used a bag-of-words model, which only takes into account individual words and their associated scores; no n-grams, full sentences, or any other type of multiword combinations are considered or incorporated in the tool. In addition, the tool was upgraded to allow further analysis by incorporating not only the previous two-day sequential analysis but also analyses covering longer durations, reaching up to several weeks. New indices, consolidated lexicons, multivariate analysis, intrinsic embedding, temporal embedding, and state-of-the-art autoencoder analyses are included as the latest features in the tool. In summary, we created a tool that supports scoring for different lexicons in the same framework, as well as the analysis of statistical and temporal dynamics, the information-theory analysis within and among groups, and all the newly defined indices and the autoencoder analysis approach.

1) DATASET EXTRACTION

The convulsive social circumstances surrounding the political environment in Spain suggests that high volumes of traffic were present in social media before, during and after the 2019 Spanish presidential elections, providing a reasonable setting to effectively evaluate and describe the power of the proposed analytical tools with sufficient statistical significance and offering a unique testing environment for the present analysis. Therefore, using the abovementioned tools, all tweets referring to the topic of the 2019 presidential elections in Spain or to key corresponding users were extracted and carefully stored for further analysis. Key users were considered those corresponding to the first and second positions of the relevant parties, as well as the official accounts for the parties themselves (see Table 1). Relevant parties were selected in terms of citizens' awareness according to public surveys. The extracted records included all available information from the social network platform; specifically,

TABLE 1. Twitter usernames for different groups, namely, political parties, 1-level politicians, and 2-level politicians.

Political parties	1-level politicians	2-level politicians
@CiudadanosCs	@Albert_Rivera	@InesArrimadas
@PSOE	@SanchezCastejon	@CarmenCalvo
@ahorapodemos	@Pablo_Iglesias	@Irene_Montero
@populares	@PabloCasado	@TeoGarciaEgea
@vox_es	@Santi_Abascal	@ivanedlm
@MasPais_Es	@ierrejon	@MartaMHiguera
@eajpv	@AndoniOrtuzar	@Aitor_Esteban
@Esquerra_ERC	@GabrielRufian	@MartaRovira
@JuntsxCatBCN	@KRLS	@Elsa_Artadi
@ehbildu	@ArnaldoOtegi	@OskarMatute
@prcantabria	@RevillaMiguelA	
	@anioramas	

we considered the following fields: user ID, username, tweet, creation date, tweet content, date and logical control of whether it was a retweet. The data were collected during the period from one month before the elections to one week after their completion to convey the different underlying dynamics for the three theoretically clearly different groups of users. Data were collected from 21 October until 22 November.

The first column of Table 1 shows the Twitter usernames of the political parties that have been considered according to their relevance to the elections. In the second and third columns, the first and second representatives of these parties are listed. For a better understanding of the results and corresponding conclusions, a brief description of the parties is presented, introducing their political perspective according to the criteria recognized by the parties themselves, as well as their geographical presence. In terms of the latter, the first five parties are represented all over the national territory, while the rest are regional parties. Among the regional parties, it is necessary to highlight a subset that considers themselves parties with ambitions of independence and self-governance (*eanpv*, *Esquerra_ERC*, *JuntsxCatBCN*, and *ehbildu*). Regarding the national parties, according to their self-recognition, they can be sorted from those closer to Communism, or left-leaning, to those closer to Conservative, or right-leaning, as follows: *ahorapodemos*, *PSOE*, *CiudadanosCs*, *populares*, *vox_es*. For a full understanding of Table 1, special mention should be given to the last first representative, Ana Oroamas (*anioramas*), who is the head of the regional party Coalición Canaria. This user was included in the study but not the party due to the very limited social activity of the corresponding Twitter account.

2) PREPROCESSING

After the selected records were extracted, the corresponding preprocessing was performed with the following characteristics: (i) Stop-Words. Stop words were removed from all tweets prior to any characterization. (ii) Symbols. Iconic, image-type symbols (emoji) were excluded during

postprocessing, but character-convertible icons were decoded for further analysis and considered simple words later. (iii) n-grams. No word grouping (n-grams), multiwords, full sentence or structured natural language processing was included in the preprocessing stage, leaving only space for later word-by-word scoring.

3) SENTIMENT SCORING STRATEGY

Although scoring could be considered effectively part of formal processing and will be covered later in the experiment section, here, we present the basic operational scope of this scoring strategy, as it was applied seamlessly and extensively throughout the experiments. Sentiment scoring was performed based on a lexicon sentiment-mapping strategy by computing the sentiment of each tweet as the aggregated value of all individual word scores according to the applied lexicon.

IV. EXPERIMENTS AND RESULTS

In this section, we address the different experiments carried out in this work. First, we benchmarked the different sentiment dictionaries available in the Spanish language as well as their descriptive capacity. Second, we analyzed the temporal dynamics throughout the aforementioned election period for the three groups of users. Third, the results of the analyses using intrinsic embeddings based on autoencoders are presented. Finally, we analyzed the temporal embeddings of the sentiments of user groups with different natures. For illustrative purposes, we incorporated here Fig. 1, which represents the full process followed in every experimental setting developed in this paper. The process began with the introduction of keywords and/or the user to extract the corresponding tweets using the developed tool. Second, the previously mentioned preprocessing procedure was applied to every extracted tweet. Then, scoring was applied using a sentiment word-mapping strategy according to an elected lexicon to quantify/detect the sentiment. Finally, classification



FIGURE 1. Illustrative representation of the experimental process workflow.

algorithms provided descriptive results for discussion and analysis.

A. COMPARISON OF DICTIONARIES

To a large extent, the analytical power of lexicon-based methods is related to the representativeness of the sentiment expressed or, more specifically, to the number of qualified words and their quantification. For that reason, we evaluated and benchmarked four sentiment lexicons widely available in the Spanish language, hereafter referred to as AFINN [56], JAEN [57], Linguakit [58], and SBU [59]. The aim of this comparison was to analyze performance and sensitivity in terms of statistical and descriptive capabilities as defined in previous works [3]. Different dictionaries can result in different characterizations of sentiment for the parameters under study due to differences in their valuation ranges and the words effectively considered. To contend with these different valuation ranges, they were normalized from -5 to 5 for all dictionaries. To further extend the predictive and expressive capacity of the different dictionaries considered in this work, a list of symbols recognized in social networks, such as emojis, were added to all of them prior to the comparisons conducted herein.

The results obtained show a wide variety of differences in terms of sensitivity, and characterization, which were very dependent on the dictionary applied. Sensitivity here refers to the descriptive capacity of the model generated with a certain lexicon that provides a larger effective intensity as a consequence of the existence of a higher number of qualified terms that undoubtedly would eventually lead to a better qualification and measurement of the evoked sentiment under a consolidated statistical perspective. Fig. 2 represents the intensity of the number of tweets qualifying for evaluation, showing relatively higher values for JAEN and Linguakit. This result is consistent with the fact that these two dictionaries include a significantly larger number of terms in their respective datasets than the other dictionaries. On the other hand, for the analysis using the rest of the dictionaries, a significant number of the evaluated tweets did not qualify for classification, as they did not include any words matching the record; therefore, they were excluded from the analysis. This higher representation of lexicon records is especially noticeable for the JAEN dictionary and the username @sanchezcastejon, hitting the maximum of 6,000 tweets, whereas AFINN barely hit 3,000 tweets, SBU 4,000 tweets, and Linguakit 5,000 tweets.

This multilexicon analysis was also conducted to analyze the sentiment histograms, entropy, and each of the previously

TABLE 2. Term coincidence among the dictionaries considered in this study (AFINN, JAEN, Linguakit, and SBU). Considering the set of unique terms jointly provided by all the dictionaries, about 70% of them (9,174, in bold font) are present in only one of the dictionaries, but not the others.

	Number of Dictionaries			
	4	3	2	1
Terms shared among dictionaries	594	1,180	2,223	9,174

defined statistics. In particular, for the sentiment histograms per user in a visual exploratory analysis, AFINN stands out with the highest neutral main lobe and an almost continuous leakage extension to either side, including limited extension to the secondary lobes. Different results were obtained with the remaining dictionaries, where one or two prominent separated secondary lobes are clearly visible. To our understanding, these outstanding secondary lobes very much relate to the almost categorical and dichotomous, not granular, quantification of the lexicons. This result is also consistent with previous considerations stated with regard to intensity; hence, it was possible to appreciate the importance of users such as @sanchezcastejon and @SantiABASCAL as clear leaders in terms of social network activity.

In terms of entropy, although a number of similarities were found among the studied dictionaries, it was difficult to generalize any one behavior across all of them.

Additionally, from a critical, visual inspection perspective, the higher granularity of AFINN does not add clear benefits, which could be justified based on the very limited number of words included within; furthermore, taking into account that the number of words played a significant role in the number of tweets that could be expressed in this analysis, we considered extending the experiments in this work either with JAEN (the largest dictionary) or building a new dictionary by consolidating all the available dictionaries to extend the sensibility potential and assessment power of the evaluated modes. To evaluate the most suitable scenario, we investigated the different terms included in the different dictionaries, identifying those shared among them (see Table 2). We found that from a total of 13,171 different terms jointly considered in the four datasets, 70% of the total (9,174 words) are only represented in one single dictionary. Hence, to extend the sensitivity beyond any individual dictionary, a combined dictionary (newLEX) was proposed for this analysis. This new lexicon was formed by previously transforming an 11-level AFINN dictionary into a dichotomic model for seamless consolidation, converting positive and negative terms to their corresponding new values.

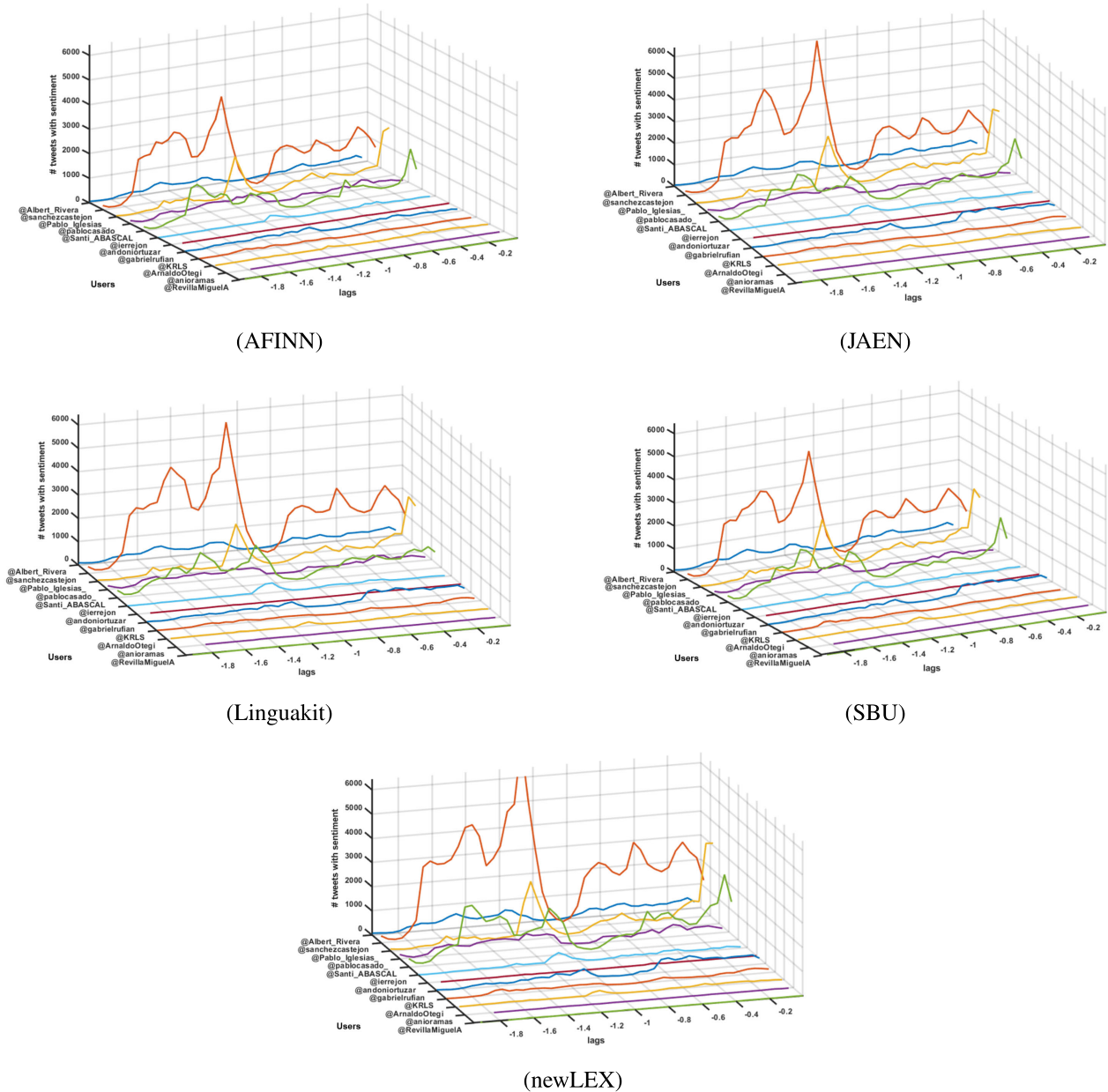


FIGURE 2. Time representations of messages from 21 and 22 October for the first-level politicians in the included parties. The plots represent the number of tweets within one-hour time windows in the dictionaries under study, namely, AFINN, JAEN, Linguakit, SBU, and newLEX.

TABLE 3. Number of terms (diagonal) and number of pairs of shared terms (superdiagonals) for each considered dictionary. Last row and column correspond to the consolidated dictionary newLEX, which included all the non-repeated terms.

	AFINN	JAEN	Linguakit	SBU	newLEX
AFINN	1,934	1,070	842	1,036	1,934
JAEN		8,222	1,876	2,746	8,222
Linguakit			5,021	1,757	5,021
SBU				4,359	4,359
newLEX					13,171

In that regard, as depicted in Table 3, JAEN was the dictionary with the largest number of terms (8,222), followed by Linguakit with 5,021 and SBU with 4,360. AFINN, with

close to two thousand terms (1,934), closes this benchmark. It is important to note at this point that previous dictionaries defined a three-level quantification (positive, negative and neutral), whereas AFINN included a much more detailed quantification proving a granularity of eleven different levels (five positive, five negative, and neutral). Together, the four dictionaries combined offered 13,171 different terms, where only 594 were effectively considered in all of them, 1,180 in three different dictionaries, and 2,223 were incorporated in two. For a further benchmark and to define the correct strategy going forward, it is important to know that the bulk of terms (9,174) were incorporated only in one single dictionary.

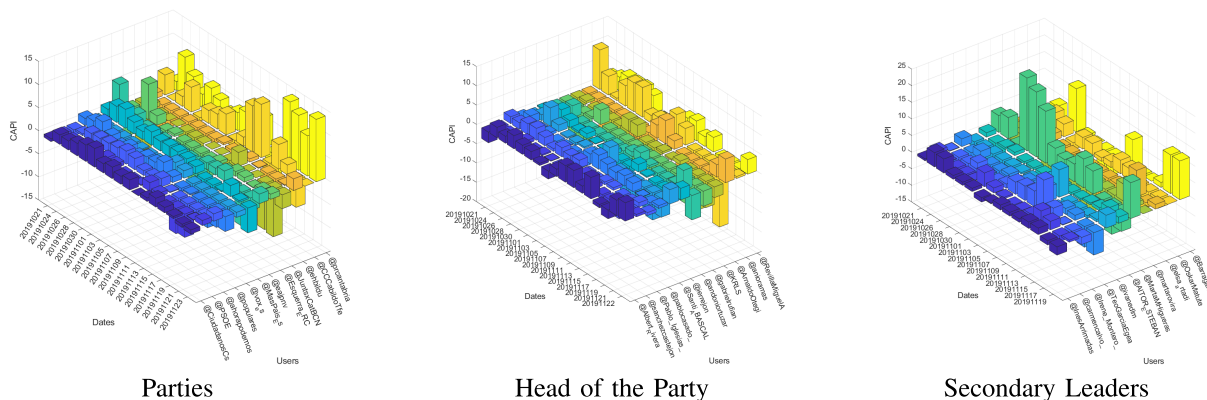


FIGURE 3. CAPI analysis throughout the observation period.

As a general result, we can state that the difficulty in finding homogeneity across the dictionaries in the different analyses suggested a relevant bias between the results and the dictionary used. We also found that the higher granularity in certain dictionaries did not necessarily relate to a higher descriptive capability. Additionally, the absence of a broad consensus in terms of the words simultaneously present in the different dictionaries suggested the need to expand the representative capabilities of the lexicons. For that reason, a new consolidated lexicon was proposed that included the terms contained within all previously presented dictionaries, following a dichotomist approach for homogeneity among terms added to the same dataset. This resulted in the creation of a new lexicon named newLEX; although we theoretically reduced the granularity for a number of terms, we increased and systematically enhanced the representative and descriptive sentiment capabilities among its previously presented peers, and for that reason, it was proposed for further development in this work.

B. TEMPORAL ANALYSIS

The distribution of tweets in cumulative displays such as histograms cannot be readily used to analyze the evolution of sentiment during the time period under observation because the number of tweets is highly variable, making comparisons difficult. Instead, we approached temporal analysis from the perspective of the CAPI, introduced in Section III-A. Since the CAPI is independent of the number of tweets, it can offer a consistent outlook on the evolution of sentiment over time.

We applied CAPI analysis to the three groups in the dataset, namely, political parties and first-level and second-level politicians for each party. Fig. 3 shows the temporal evolution of the index across these three entities. The results show that the temporal evolution of sentiment does not follow parallel profiles across parties or first- or second-level party politicians. Notable variability was found in all cases but in different directions. Among the regional parties, as defined in the previous section, we observed a much higher variability in those parties advocating for independence. Those political parties self-described as firm in their ideas at either limit of

the political spectrum (left or right) also showed significantly higher variability in the observation period.

It can also be observed that the leaders of independence-advocating parties showed a decrease in their positivity indices toward the end of the period, most notably during the week before the vote and on the second week after. During the election week, the positivity index of most party leaders remained stable, except for some of the pro-independence leaders. Regarding the second-level party politicians, higher CAPI variability was found among the representatives of regional parties, whereas those representing national parties showed more consistent and stable profiles. During the week before and of the election, greater variability was observed among the indices of the regional parties. Two weeks after the vote, however, smaller changes were observed in the CAPI.

If we examine the political parties themselves, generally, positivity was higher before and during the election but lower in the weeks following the election. One exception was seen for the regional parties, for whom this high positivity was only visible during the week of the election. Furthermore, the regional independence-advocating parties tended to demonstrate notably negative index values, especially during the weeks prior to election and, in some cases, the very same days of the elections.

Entropy analysis throughout the observation period showed uneven results with respect to variability, as displayed in Fig. 4 for the same entities. Regional parties consistently showed lower entropy values, although they also presented higher variability throughout the period. In contrast, entropy was higher for the national parties, but it remained essentially stable during the entire observation period.

Finally, we also conducted temporal analysis of relevant statistical measures, including intensity, standard deviation, mean, and their respective autocorrelations. Our observations confirm previous findings [3] in terms of the significant seasonality of the signals, with clear daily and weekly trends.

C. INTRINSIC EMBEDDINGS

As described in Section III-B, autoencoders can be used to learn a mapping from the original data space to a

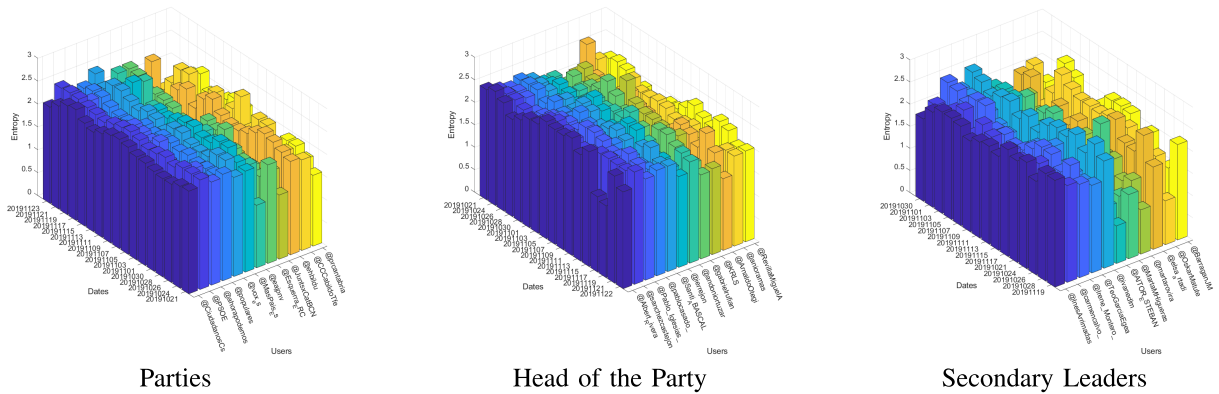


FIGURE 4. Entropy analysis throughout the observation period.

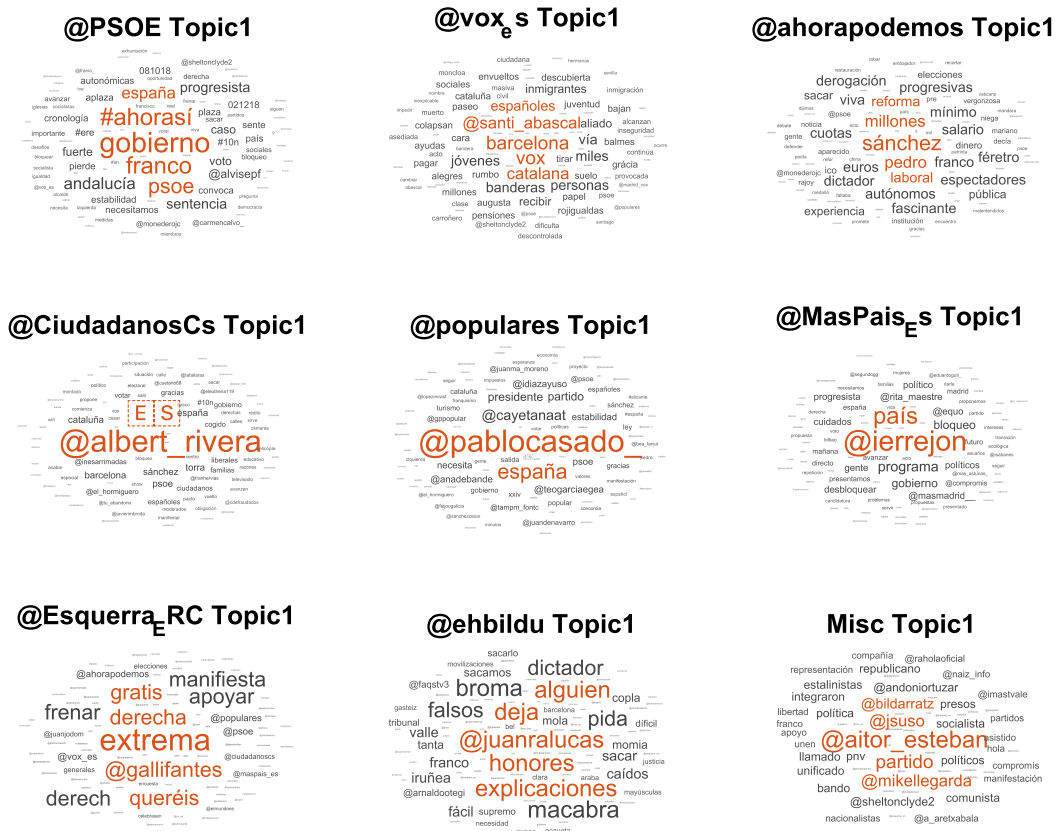


FIGURE 5. Word clouds representing the LDA fitting of topics per political group.

lower-dimensional space. In the experiments described in this section, we created autoencoders from a bag-of-words model (extracted from the tweets) to a three-dimensional latent space, which allowed for visual interpretation without resorting to further projections or transformations.

Due to the heavy reduction in dimensionality –from thousands of features to just 3 latent dimensions– we restricted this experiment to a tiny subset of the dataset. Our focus was on the interpretability and visualization of results rather than on prediction accuracy.

These experiments were thus performed on a dataset of tweets related to the main political parties and collected

on 30 and 31 October 2019. This subset was presumed to be representative of the overall state of mind at a point in time sufficiently close to the election but not so close as to be distorted from the increased polarization typical of this period. Tweets were extracted from the dataset together with their sentiment scores according to the newLEX dictionary described in Section IV-A.

Since the distribution of tweets among parties was asymmetrical, we heuristically defined a threshold of 2,000 documents, combining tweets from parties below that threshold in a common miscellaneous group. This resulted in a total of 9 distinct political groups instead of the original

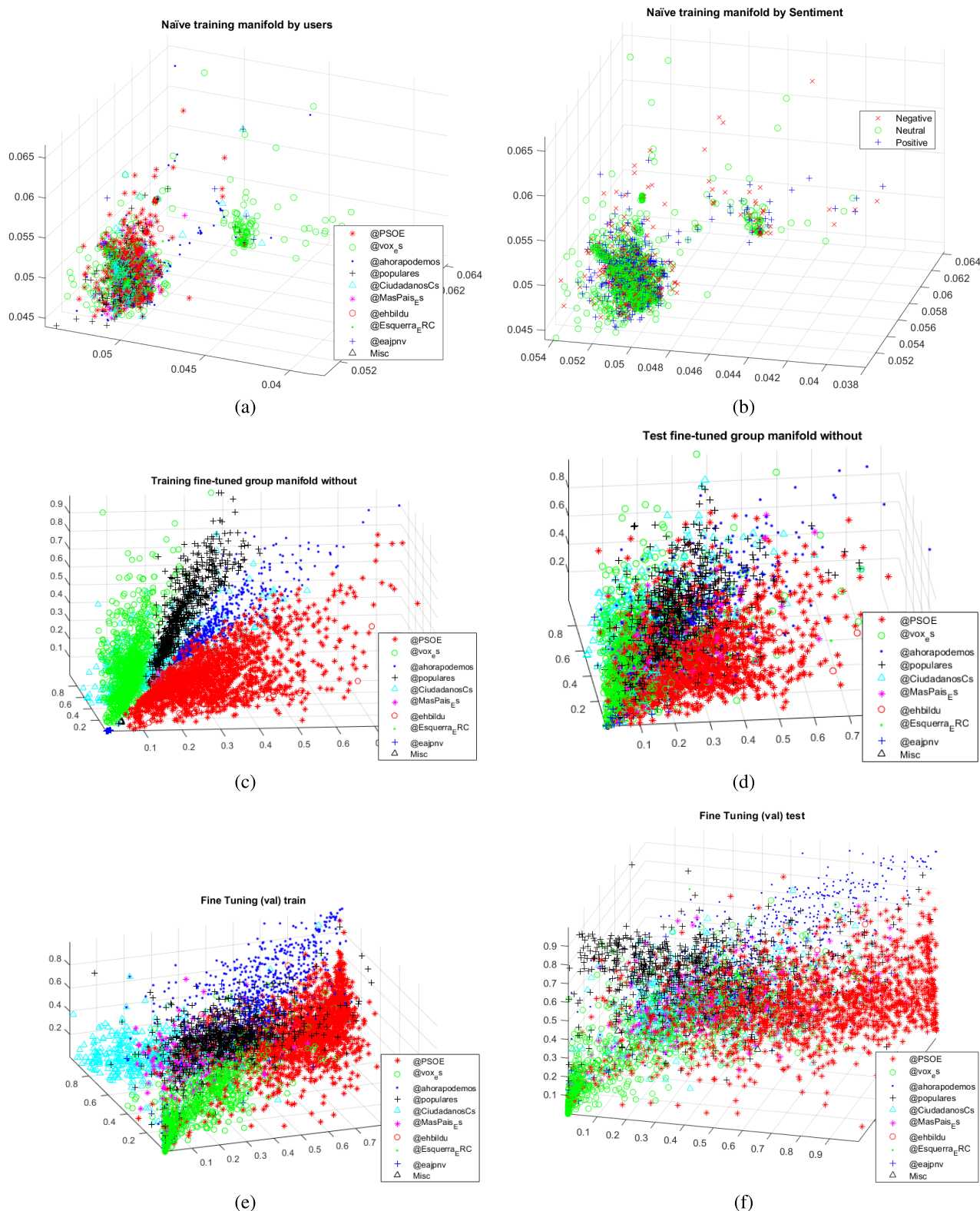


FIGURE 6. Representation of autoencoder-generated embeddings. Top row: Naïve autoencoder with 3 latent dimensions. Middle row: Encoder after fine-tuning to party classification. Bottom row: Fine-tuned results with validation and early stopping.

11 parties, where tweets from the @eajpnv, @JuntsxCatBCN, and @prcantabria parties were all assigned to the misc group. Coincidentally, these three parties have a regionalist profile, representing different regions of the state.

Documents were created by tokenizing all tweets but excluding mentions of the Twitter handle belonging to the corresponding political party. This was done to prevent trivial solutions during the fine-tuning phase, where any tweet

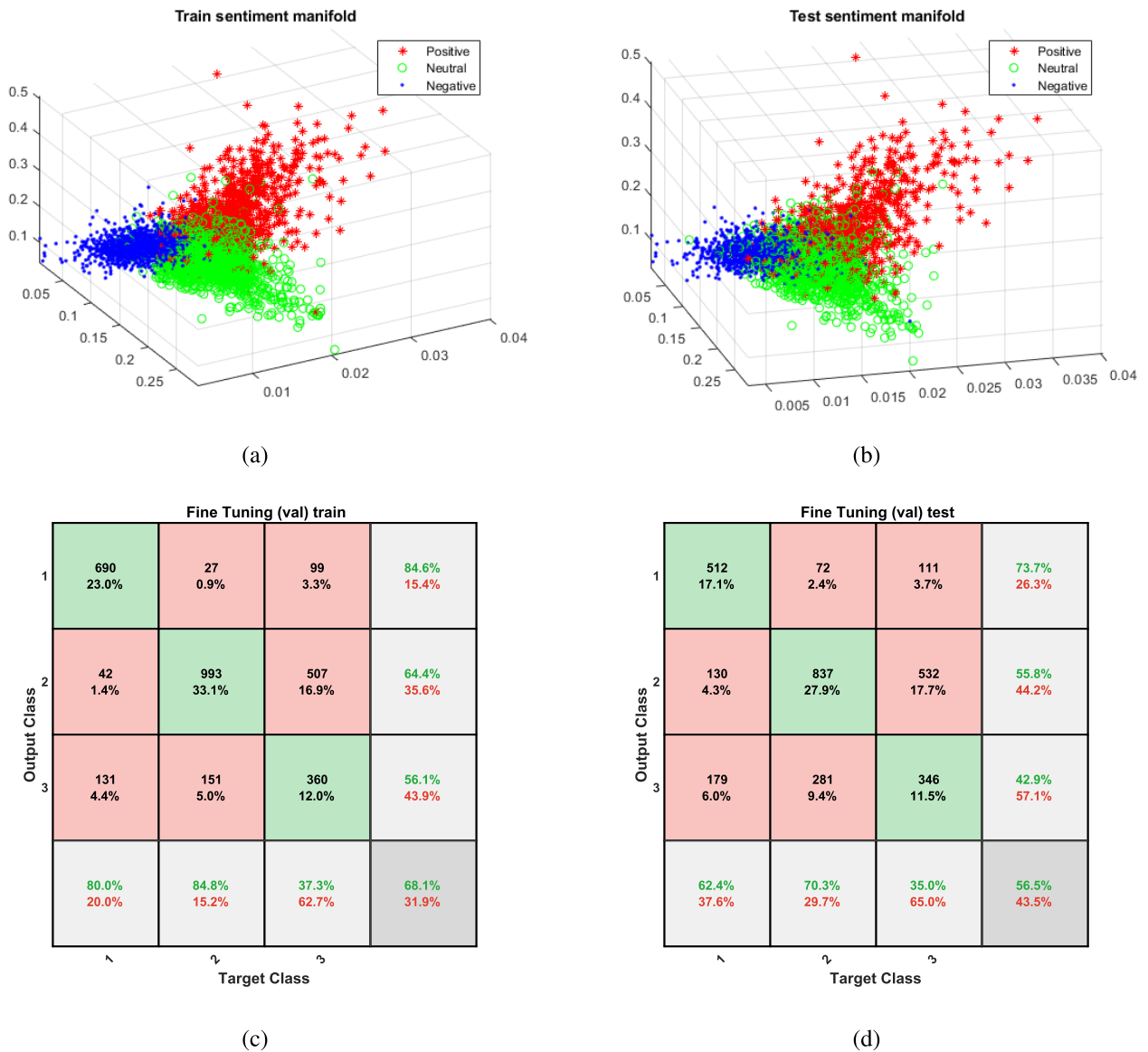


FIGURE 7. Fine-tuned encoder used to learn sentiment classification. Top row: Visual representation. Bottom row: Confusion matrices.

containing the Twitter handle of a group would be assigned to that group with high probability. A bag-of-words model was extracted from the union of all documents across all groups. In addition, per-group bag-of-words models were also created and used to perform latent Dirichlet allocation (LDA) for two topics, the result of which is shown in Fig. 5. The figure shows that messages related to a few parties were strongly centered around their leaders – this occurred across the entire political spectrum, in right-leaning (*@populares*), left-leaning (*@MasPaisEs*) and centrist (*@CiudadanosCs*) parties. Other topics focus on keywords that were important in the context of those parties, and references to other parties were also frequent. For example, left-wing and Republican party *@EsquerraERC* tweets frequently referred (perhaps worryingly) to the “far-right” topic. These per-group LDA

results were not used in the rest of the experiments, but they seem to support the notion that some homogeneity can be inferred from the documents in each group.

Training and test subsets were created by randomly sampling documents from the dataset. Both subsets were built with 3,000 tweets each. A naïve autoencoder was then trained on the training subset, configured to learn embeddings on a three-dimensional latent space. The top row of Fig. 6 shows the results of this training procedure. Panel (a) depicts the embedding of the test set to the latent space, where each dot represents a document, colored according to the group to which they belong. Some groupings are visible, but there is no clear structure to the segmentation. Panel (b) shows the same documents but colored according to their sentiment scores as gathered during the collection phase and categorized into

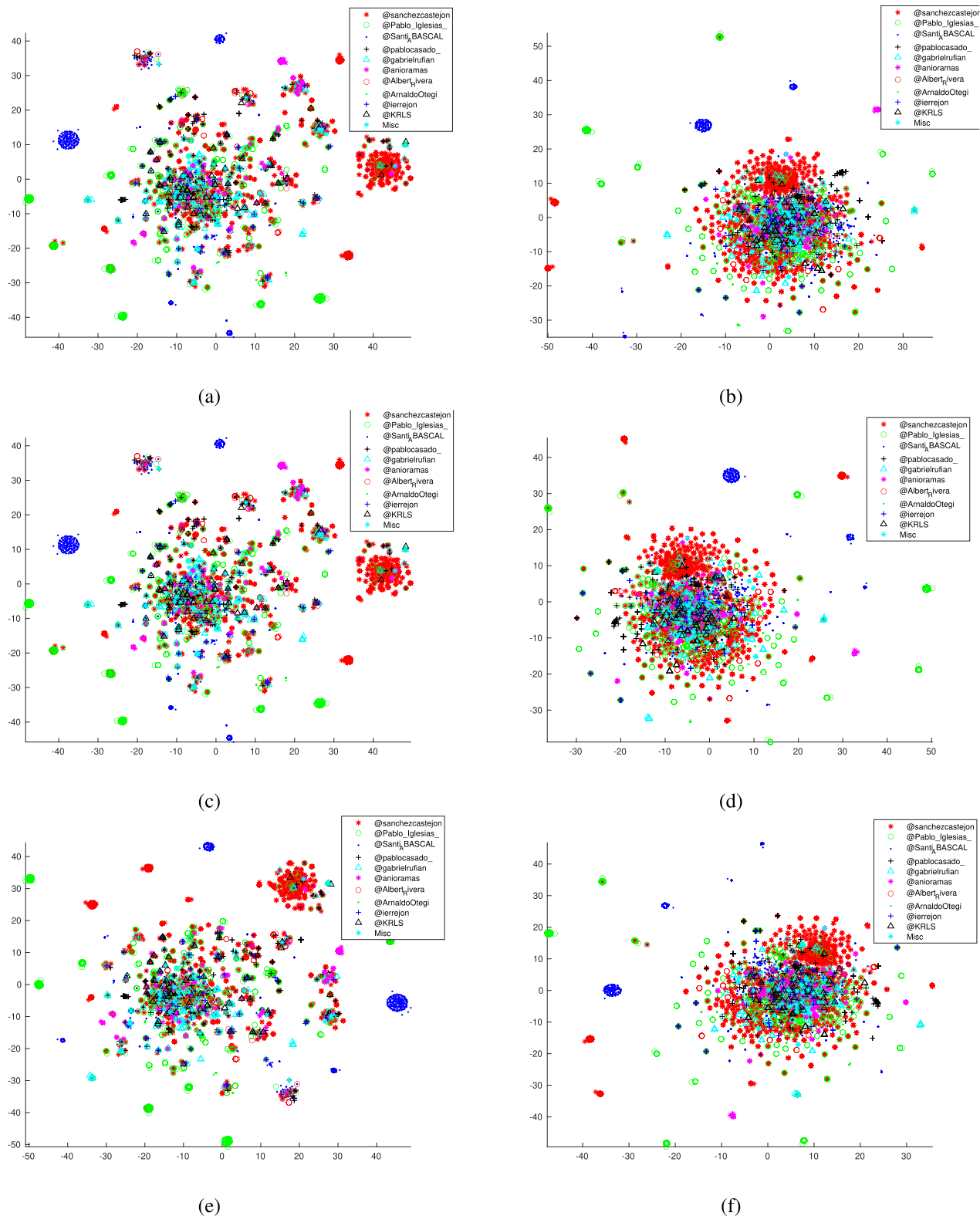


FIGURE 8. Temporal sentiment embedding using tSNE.

three classes, namely, negative, neutral, and positive. Again, the separability in the latent space appears to be poor.

The naïve autoencoder was then fine-tuned by removing the decoder component of the network and appending a

softmax layer, training the system to learn the classification of each document according to the group from which it was taken. Training was performed on the training set for a fixed number of epochs. The middle row in Fig. 6 shows

the representation of the embeddings obtained for both the training and test sets. The training set in panel (a) shows a very clear separation among groups; the test set in panel (b), however, shows a lack of generalization, as demonstrated by the loss of precision, even if the groups are still located in the same regions of the 3D latent space.

A second fine-tuning training process was then performed. The setup was the same as in the previous experiment, but a validation set was extracted from the training set (20% of the samples), and the process was stopped early when the validation error stopped improving. The condition was met very quickly, and the results are depicted in the bottom row of Fig. 6. Again, panel (a) shows the training set, and panel (b) shows the test set. The groups are now spread across larger areas, and there is some overlap among them, but the areas are better organized according to the political spectrum. It is also clear that the generalization performance of this model has been improved.

The following regularization hyperparameters were considered for this experiment: L2 weight regularization, sparsity proportion, and sparsity regularization based on Kullback-Leibler divergence. They all had a minor impact in the convergence process, and the groupings observed were also similar. The bottom row of Fig. 6 was obtained using a L2 weight regularization factor of 0.001, a sparsity regularization coefficient of 1, and a sparsity proportion of 0.05.

Finally, a sentiment classifier was created based on the encoder portion of the fine-tuned model with validation. Fig. 7 shows how, in this case, there is separation between the three classes, and the results are similar for both the training and test sets. Tweets with positive sentiment appear to be more easily classified than others, and neutral tweets have the most patent overlap with other classes.

D. SENTIMENT TEMPORAL EMBEDDINGS

Finally, we used tSNE to visualize high-dimensional data in a low-dimensional embedding. Considering the high dimensionality of the input space, we proposed two strategies to achieve better visualizations. In the first, we considered all features, whereas in the second, we leveraged the potential of PCA to reduce the number of dimensions to a reasonable amount (fixed here to 50). This is especially recommended when the number of features is very high.

As previously described, tweets related to the main political leaders and their parties were collected and analyzed starting a month before the presidential elections in Spain and up to one week after they were held; see Fig. 8 for details. In panel (a), we represent the embedding obtained without considering PCA, whereas panel (b) shows the embedding obtained with PCA. The figure shows that some tweets are grouped together and separated from the others. After a thorough analysis of these tweets, we conclude that they are retweets associated with some controversial topic posted in a specific period of time. Additionally, to analyze the obtained embeddings from a temporal viewpoint, we scrutinized the data associated with the first week

individually, in panels (c) and (d), and with the last week, in panels (e) and (f), for which similar conclusions are obtained.

V. DISCUSSION AND CONCLUSION

In this study, we aimed to analyze the temporal evolution of sentiment measurements in an intense scenario involving political elections. Whereas the majority of works in the literature lately have focused on the use of machine learning engines, we chose to perform knowledge extraction using both specified indices based on fundamental magnitudes from information theory and intrinsic features obtained by nonlinear mappings to low-dimensional embeddings. This approach aims to offer advantages in terms of interpretability and the evolution of temporal dynamics.

Our results on the relevance of the lexicons used and the need for their consolidation prior to any analysis are consistent with the conclusions by others in this setting [59], [60]. The simple translation of lexicons with sentiment from other languages (mostly English) is not necessarily a good-quality approach, and special attention has to be paid to this basic step. Greater attention should be devoted to the use and development of lexicons; for example, they could be extended or combined as usual in the field with existing word embeddings.

Consistent with other previous studies [3], seasonality in sentiment is strongly dependent on the number of messages, but the behavior and dynamics are different when controlling for this factor. Explicit indices and nonlinearly extracted features can also be consistent and complementary in terms of knowledge discovery. On the one hand, our analysis of entropy and mutual information showed that different dynamics can be observed for different parties and leaders and that they often follow different paths. The changes in the CAPI are also consistent with the dynamics highlighted by the information theory indices and can offer a less inconclusive image of the temporal dynamics. On the other hand, low-dimensional embeddings from naïve autoencoders can be used to extract and show social patterns, but caution has to be taken in terms of possible overfitting, and moreover, their embedding patterns are more prone to reality and are better defined when fine-tuned with certain criteria. For instance, the expected distance relationship among left-leaning, right-leaning, and other parties is more consistent with our expected experience after fine-tuning, which should be taken into account when addressing this kind of problem. Cross-sectionally built static embeddings are able to reflect changes over time, such as the changes in positiveness from early-to- to near-election periods; furthermore, specific clusters can reflect their entity when associated with controversial facts, which seem to present with their own dynamics through this embedding view.

Political results show that both indices and manifold features provide informative characterizations of the sentiment dynamics in terms of temporal behavior and defined semantic domains, which were used to subsequently draw the

participants' sentiment spaces, with them being main actors or simply users. In addition, variations in sentiment behavior were found across political parties and main players, especially when they were analyzed according to whether the political party they represented was national, regional, centrist or separatist. We believe these results can be explained by the varying number of tweets that parties posted and the language used by Twitter users. The regional parties received fewer tweets than the national parties, and some of their users wrote in other languages, such as Catalan or Basque. Greater polarity was found in the sentiment indices and features of regional separatists during the weeks prior to elections than among the members of other national and regional parties.

According to the positivity index, we found that the frontline representatives of regional separatist parties showed a higher index of negativity than the representatives of regional nonseparatist parties, who, in turn, showed a high index of positivity. National political parties, meanwhile, were situated at approximately 0 or a negative index. Therefore, in summary, the index of positivity showed greater variability among the frontline representatives of regional separatist parties than among the representatives of nationalist parties, where more stability was inferred, except among parties with more extreme political tendencies, where we again found variability.

In this work, we scrutinized the usefulness of indices with moderate complexity to obtain information on sentiment in politics, focusing on interpretability and temporal dynamics. We conclude that both specific measurements (from information theory or experience principles) and non-linearly extracted features through embeddings can provide us with the ability to perform knowledge extraction in this setting. Different from other works, we did not focus here on advanced deep architectures, but these should be taken into account because of the high power they currently show in this setting. Interpretability and temporal dynamics are two desirable features to incorporate into more advanced schemes in sentiment analysis technology.

The results obtained with the mentioned strategies pave the way for new developments and investigations that delve into the same or larger data sets, as well as offers an opportunity to validate or benchmark over different disciplines and new datasets.

REFERENCES

- [1] A. Kumar and A. Jaiswal, "Systematic literature review of sentiment analysis on Twitter using soft computing techniques," *Concurrency Comput., Pract. Exper.*, vol. 32, no. 1, p. e5107, Jan. 2020.
- [2] C. Lynch, C. O'Leary, G. Smith, R. Bain, J. Kehoe, A. Vakaloudis, and R. Linger, "A review of open-source machine learning algorithms for Twitter text sentiment analysis and image classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–9.
- [3] M. Rodríguez-Ibáñez, F.-J. Gimeno-Blanes, P. M. Cuenca-Jiménez, S. Muñoz-Romero, C. Soguero, and J. L. Rojo-Álvarez, "On the statistical and temporal dynamics of sentiment analysis," *IEEE Access*, vol. 8, pp. 87994–88013, 2020.
- [4] M. P. Anto, M. Antony, K. M. Muhsina, N. Johny, V. James, and A. Wilson, "Product rating using sentiment analysis," in *Proc. Int. Conf. Elect., Electron., Optim. Techn.*, Mar. 2016, pp. 3458–3462.
- [5] M. Z. Ansari, M. B. Aziz, M. O. Siddiqui, H. Mehra, and K. P. Singh, "Analysis of political sentiment orientations on Twitter," *Procedia Comput. Sci.*, vol. 167, pp. 1821–1828, Jan. 2020.
- [6] B. Bradbury and I. Kim, "What if 140 characters can impact your company's market value? Tweeting corporate social responsibility and stock prices in the financial sector," in *Proc. Surfing IT Innov. Wave- Americas Conf. Inf. Syst.* Atlanta, GA, USA: Association for Information Systems, 2016, pp. 1–10.
- [7] S. Y. Yang, S. Y. K. Mo, and A. Liu, "Twitter financial community sentiment and its predictive relationship to stock market movement," *Quant. Finance*, vol. 15, no. 10, pp. 1637–1656, Oct. 2015.
- [8] M. S. Neethu and R. Rajasree, "Sentiment analysis in Twitter using machine learning techniques," in *Proc. 4th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2013, pp. 1–5.
- [9] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *CS224N Project Rep., Stanford*, vol. 1, no. 12, p. 2009, 2009.
- [10] A. Porshnev, I. Redkin, and A. Shevchenko, "Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis," in *Proc. IEEE 13th Int. Conf. Data Mining Workshops*, Dec. 2013, pp. 440–444.
- [11] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, pp. 2674–2703, 2018.
- [12] J. Lemley, S. Bazrafkan, and P. Corcoran, "Deep learning for consumer devices and services: Pushing the limits for machine learning, artificial intelligence, and computer vision," *IEEE Consum. Electron. Mag.*, vol. 6, no. 2, pp. 48–56, Apr. 2017.
- [13] Z. Obermeyer and E. J. Emanuel, "Predicting the future—Big data, machine learning, and clinical medicine," *New England J. Med.*, vol. 375, no. 13, pp. 1216–1219, 2016.
- [14] K. Jindal and R. Aron, "A systematic study of sentiment analysis for social media data," *Materials Today: Proceedings*, Feb. 2021.
- [15] A. Khatua, A. Khatua, and E. Cambria, "Predicting political sentiments of voters from Twitter in multi-party contexts," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106743.
- [16] S. Rill, D. Reinel, J. Scheidt, and R. V. Zicari, "PoliTwi: Early detection of emerging political topics on Twitter and the impact on concept-level sentiment analysis," *Knowl.-Based Syst.*, vol. 69, pp. 24–33, Oct. 2014.
- [17] D. Liu and L. Lei, "The appeal to political sentiment: An analysis of donald Trump's and hillary Clinton's speech themes and discourse strategies in the 2016 US presidential election," *Discourse, Context Media*, vol. 25, pp. 143–152, Oct. 2018.
- [18] E. Kušen and M. Strembeck, "Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 austrian presidential elections," *Online Social Neww. Media*, vol. 5, pp. 37–50, Mar. 2018.
- [19] B. Joyce and J. Deng, "Sentiment analysis of tweets for the 2016 US presidential election," in *Proc. IEEE MIT Undergraduate Res. Technol. Conf. (URTC)*, Nov. 2017, pp. 1–4.
- [20] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. 8th Int. AAAI Conf. Weblogs Social Media*, 2014, pp. 1–10.
- [21] A. Makazhanov and D. Rafiei, "Predicting political preference of Twitter users," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2013, pp. 298–305.
- [22] N. A. Diakopoulos and D. A. Shamma, "Characterizing debate performance via aggregated Twitter sentiment," in *Proc. 28th Int. Conf. Hum. Factors Comput. Syst. (CHI)*, 2010, pp. 1195–1198.
- [23] R. Bose, R. K. Dey, S. Roy, and D. Sarddar, "Analyzing political sentiment using Twitter data," in *Information and Communication Technology for Intelligent Systems*, vol. 107. Singapore: Springer, 2019, pp. 427–436.
- [24] R. Castro, L. Kuffó, and C. Vaca, "Back to #6D: Predicting venezuelan states political election results through Twitter," in *Proc. 4th Int. Conf. eDemocracy eGovernment (ICEDEG)*, Apr. 2017, pp. 148–153.
- [25] D. Watts, K. M. George, T. K. A. Kumar, and Z. Arora, "Tweet sentiment as proxy for political campaign momentum," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 2475–2484.
- [26] M. Campanale and E. G. Caldarola, "Revealing political sentiment with Twitter: The case study of the 2016 Italian constitutional referendum," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2018, pp. 861–868.
- [27] J. S. Cesteros, A. Almeida, and D. L. De Ipiña, "Sentiment analysis and polarity classification in Spanish tweets," in *Proc. Workshop Sentiment Anal. Int. Conf. Spanish Soc. Natural Lang. Process.*, 2015, pp. 23–28.

- [28] L. Oikonomou and C. Tjortjij, "A method for predicting the winner of the USA presidential elections using data extracted from Twitter," in *Proc. South-Eastern Eur. Design Autom., Comput. Eng., Comput. Netw. Soc. Media Conf. (SEEDA CECNSM)*, Sep. 2018, pp. 1–8.
- [29] M. Z. Ansari, M. B. Aziz, M. O. Siddiqui, H. Mehra, and K. P. Singh, "Analysis of political sentiment orientations on Twitter," *Procedia Comput. Sci.*, vol. 167, pp. 1821–1828, Jan. 2020.
- [30] D. G. Avello, P. Metaxas, and E. Mustafaraj, "Limits of electoral predictions using Twitter," in *Proc. 5th Int. Assoc. Advancement Artif. Intell.*, 2011.
- [31] J. A. Cerón-Guzmán and E. León-Guzmán, "A sentiment analysis system of spanish tweets and its application in colombia 2014 presidential election," in *Proc. IEEE Int. Conferencas Big Data Cloud Comput. (BDCloud), Social Comput. Netw. (SocialCom), Sustain. Comput. Commun. (Sustain-Com) (BDCloud-SocialCom-SustainCom)*, Oct. 2016, pp. 250–257.
- [32] M. Abdullah and M. Hadzikadic, "Sentiment analysis of Twitter data: Emotions revealed regarding donald trump during the 2015-16 primary debates," in *Proc. IEEE 29th Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2017, pp. 760–764.
- [33] F. Nausheen and S. H. Begum, "Sentiment analysis to predict election results using Python," in *Proc. 2nd Int. Conf. Inventive Syst. Control (ICISC)*, Jan. 2018, pp. 1259–1262.
- [34] J. N. Franco-Riquelme, A. Bello-García, and J. Ordieres-Meré, "Indicator proposal for measuring regional political support for the electoral process on Twitter: The case of Spain's 2015 and 2016 general elections," *IEEE Access*, vol. 7, pp. 62545–62560, 2019.
- [35] B. A. Marina, "How people attach sentiment to emoticons and words in tweets," in *Proc. SocialCom Int. Conf.*, 2013.
- [36] L. Recalde, J. Mendieta, L. Boratto, L. Teran, C. Vaca, and G. Baquerizo, "Who you should not follow: Extracting word embeddings from tweets to identify groups of interest and hijackers in demonstrations," *IEEE Trans. Emerg. Topics Comput.*, vol. 7, no. 2, pp. 206–217, Apr. 2019.
- [37] B. Heredia, J. D. Prusa, and T. M. Khoshgoftaar, "The impact of malicious accounts on political tweet sentiment," in *Proc. IEEE 4th Int. Conf. Collaboration Internet Comput. (CIC)*, Oct. 2018, pp. 197–202.
- [38] N. N. Alnajran, K. A. Crockett, D. McLean, and A. Latham, "A word embedding model learned from political tweets," in *Proc. 13th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2018, pp. 177–183.
- [39] L. Belcastro, R. Cantini, F. Marozzo, D. Talia, and P. Trunfio, "Learning political polarization on social media using neural networks," *IEEE Access*, vol. 8, pp. 47177–47187, 2020.
- [40] L. Belcastro, F. Marozzo, D. Talia, and P. Trunfio, "Developing a cloud-based algorithm for analyzing the polarization of social media users," in *Proc. Int. Symp. Algorithmic Aspects Cloud Comput.*, Sep. 2019, pp. 15–24.
- [41] D. Antonakaki, D. Spiliotopoulos, C. V. Samaras, S. Ioannidis, and P. Fragopoulou, "Investigating the complete corpus of referendum and elections tweets," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2016, pp. 100–105.
- [42] G. Di Fatta, J. J. Reade, S. Jaworska, and A. Nanda, "Big social data and political sentiment: The tweet stream during the UK general election 2015 campaign," in *Proc. IEEE Int. Conf. Smart City/SocialCom/SustainCom (SmartCity)*, Dec. 2015, pp. 293–298.
- [43] X. Zhou, X. Tao, J. Yong, and Z. Yang, "Sentiment analysis on tweets for social events," in *Proc. IEEE 17th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, Jun. 2013, pp. 557–562.
- [44] T. Seçkin and Z. H. Kilimci, "The evaluation of 5G technology from sentiment analysis perspective in Twitter," in *Proc. Innov. Intell. Syst. Appl. Conf. (ASYU)*, Oct. 2020, pp. 1–6.
- [45] S. Paliwal, S. K. Khatri, and M. Sharma, "Sentiment analysis and prediction using neural networks," in *Proc. Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Jul. 2018, pp. 1035–1042.
- [46] L. Zhang, "The research on general case-based reasoning method based on TF-IDF," in *Proc. 2nd Int. Conf. Saf. Produce Informatization (IICSPI)*, Nov. 2019, pp. 670–673.
- [47] P. Berka, "Sentiment analysis using rule-based and case-based reasoning," *J. Intell. Inf. Syst.*, vol. 55, no. 1, pp. 51–66, Aug. 2020.
- [48] C. N. dos Santos, I. Melnyk, and I. Padhi, "Fighting offensive language on social media with unsupervised text style transfer," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2018, pp. 189–194.
- [49] L. Zhang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment analysis for e-commerce product reviews in chinese based on sentiment lexicon and deep learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020.
- [50] K. Z. Aung and N. N. Myo, "Sentiment analysis of students' comment using lexicon based approach," in *Proc. IEEE/ACIS 16th Int. Conf. Comput. Inf. Sci. (ICIS)*, May 2017, pp. 149–154.
- [51] H. Zhang, W. Gan, and B. Jiang, "Machine learning and lexicon based methods for sentiment classification: A survey," in *Proc. 11th Web Inf. Syst. Appl. Conf.*, Sep. 2014.
- [52] J. P. Cunningham and Z. Ghahramani, "Linear dimensionality reduction: Survey, insights, and generalizations," *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 2859–2900, 2015.
- [53] J. A. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction*. Springer, 2007.
- [54] L. Van Der Maaten, E. Postma, and J. Van den Herik, "Dimensionality reduction: A comparative," *J. Mach. Learn. Res.*, vol. 10, nos. 66–71, p. 13, 2009.
- [55] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proc. ICML Workshop Unsupervised Transf. Learn.*, 2012, pp. 37–49.
- [56] F. A. A. Nielsen, "A new: Evaluation of a word list for sentiment analysis in microblogs," in *Proc. Workshop 'Making Sense Microposts', Big Things Come Small Packages (ESWC)*, vol. 718, May 2011, pp. 93–98.
- [57] E. Martínez-Cámara, M. T. Martín-Valdivia, M. D. Molina-González, and L. A. U. Lopez, "Bilingual experiments on an opinion comparable corpus," in *Proc. 4th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal.*, 2013, pp. 87–93.
- [58] C. Banea, R. Mihalcea, J. Wiebe, and S. Hassan, "Multilingual subjectivity analysis using machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2008, pp. 127–135.
- [59] Y. Chen and S. Skiena, "Building sentiment lexicons for all major languages," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2014, pp. 383–389.
- [60] S. M. Mohammad and D. P. Turney, "Crowdsourcing a word-emotion association lexicon," *Comput. Intell.*, vol. 29, no. 3, pp. 436–465, 2013.

• • •