Accepted Manuscript

Automatic Detection of Satire in Twitter: A psycholinguistic-based approach

María del Pilar Salas-Zárate , Mario Andrés Paredes-Valverde ,
Miguel Ángel Rodríguez-García , Rafael Valencia-García ,
Giner Alor-Hernández

PII: S0950-7051(17)30185-5
DOI: [10.1016/j.knosys.2017.04.009](https://doi.org/10.1016/j.knosys.2017.04.009)
Reference: KNOSYS 3889

To appear in: *Knowledge-Based Systems*

Received date: 25 October 2016
Revised date: 20 April 2017
Accepted date: 22 April 2017

Please cite this article as: María del Pilar Salas-Zárate , Mario Andrés Paredes-Valverde , Miguel Ángel Rodríguez-García , Rafael Valencia-García , Giner Alor-Hernández , Automatic Detection of Satire in Twitter: A psycholinguistic-based approach, *Knowledge-Based Systems* (2017), doi: [10.1016/j.knosys.2017.04.009](https://doi.org/10.1016/j.knosys.2017.04.009)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Automatic Detection of Satire in Twitter: A psycholinguistic-based approach

María del Pilar Salas-Zárate¹, Mario Andrés Paredes-Valverde¹, Miguel Ángel Rodríguez-García², Rafael Valencia-García¹, Giner Alor-Hernández³

¹Departamento de Informática y Sistemas, Universidad de Murcia. Campus de Espinardo 30100 Murcia, Spain. E-mail: {mariapilar.salas, marioandres.paredes, valencia}@um.es

²Computational Bioscience Research Center, King Abdullah University of Science and Technology, 4700 KAUST, P.O. Box 2882, Thuwal 23955-6900, Kingdom of Saudi Arabia. Email: miguel.rodriguezgarcia@kaust.edu.sa

³Division of Research and Postgraduate Studies, Instituto Tecnológico de Orizaba, Av. Oriente 9, No. 852, Col. E. Zapata, CP 94320 Orizaba, Veracruz, Mexico. E-mail: galor@itorizaba.edu.mx

Abstract

In recent years, a substantial effort has been made to develop sophisticated methods that can be used to detect figurative language, and more specifically, irony and sarcasm. There is, however, an absence of new approaches and research works that analyze satirical texts. The recognition of satire by sentiment analysis and Natural Language Processing (NLP) applications is extremely important because it can influence and change the meaning of a statement in varied and complex ways. We used this understanding as a basis to propose a method that employs a wide variety of psycholinguistic features and which detects satirical and non-satirical text. We then went on to train a set of machine learning algorithms that would enable us to classify unknown data. Finally, we conducted several experiments in order to detect the most relevant features that generate a better pattern as regards detecting satirical texts. We evaluated the effectiveness of our method by obtaining a corpus of satirical and non-satirical news from Mexican and Spanish twitter accounts. Our proposal obtained encouraging results, with an F-measure of 85.5% for Mexico and one of 84.0% for Spain. Moreover, the results of the experiment showed that there is no significant difference between Mexican and Spanish satire.

Keywords

Computational psycholinguistics, LIWC, Machine learning, Satire, Twitter.

1. Introduction

Figurative language, which consists of elements such as irony, sarcasm and satire, has been considered a pervasive phenomenon on social media platforms such as Twitter [1]. The detection of figurative language has, therefore, been fundamental for certain approaches in fields such as sentiment analysis [2], since irony-laden expressions can play the role of polarity reversers [3].

Several computational approaches whose objective is to identify sarcasm and irony have appeared in recent years [4][5][6][7][8][9]. However, satire is not often studied in literature. Satire is an important language phenomenon consisting of the use of humor and irony to criticize and ridicule someone or something. Parody, burlesque, exaggeration, juxtaposition, comparison, analogy, and double entendres are all frequently used in satirical speech and writing. News satire, also termed as fake news, is a type of parody presented in a format typical of mainstream journalism, and is called satire because of its content. News satire is particularly popular on the Web, and specifically in social networks in which it is relatively easy to mimic a credible news source and stories may achieve a wide distribution from almost any site. However, news satire is often mistaken for legitimate news, especially when it is disassociated from its original source, as occurs when a tweet from a satirical Twitter account is recommended via Facebook. Some Spanish satirical news sources are “El Deforma”, “El Jueves”, “El Dizque”, and “El Mundo Today”.

In this work, we propose a method that can be used to detect satirical texts. We therefore analyze

psycholinguistic features in order to capture not only the content but also the style of the message. We consequently used the LIWC (“Linguistic Inquiry and Word Count”) dictionary [10], which has been successfully utilized to explore the linguistic style of social media content [11]. The main motivation for the use of the LIWC is its successful application as regards figurative language. For instance, Mihalcea & Pulman [12] identified the 10 most discriminant LIWC features for humorous texts, namely You, I, swear, self, sexual, groom, cause, sleep, pronoun, and humans. Meanwhile, in Justo et al. [8] semantic LIWC features were used for the detection of sarcasm and nastiness. With regards to satire, two works have used the LIWC. Rubin et al. [13] used only two features (negative affect and punctuation), while Skalicky & Crossley [14] selected 11 features (negation, negative emotions, quantifiers, certainty, tentative, exclusion, inclusion, discrepancy, causation, past, and present). In the latter case, the features were chosen on the basis of Campbell & Katz’s [15] proposal for irony detection. Since the LIWC contains a wide set of features that make it possible to identify the differences between figurative language and literal language and many studies have selected features on theoretical grounds, we decided to analyze the complete set of LIWC features with the aim of determining which are the most discriminant for satire detection.

We then used the psychological and linguistic features to train three machine learning algorithms, namely J48, SMO, and BayesNet. Finally, a set of experiments was carried out in order to evaluate the performance of our method. These experiments were performed on a news corpus, which was collected beforehand from four satirical news accounts and four non-satirical news accounts.

It is important to mention that most of the studies on figurative language as regards addressing satire detection found deal only with the English language, perhaps owing to the lack of resources in other languages. It is for this reason that this work is principally focused on the Spanish language. Spanish is, according to the Internet World State rank¹, currently the third most used language on the Internet, and we therefore firmly believe that the computerization of Internet domains in this language is of the utmost importance.

The remainder of this paper is organized as follows. Section 2 presents a review of the literature concerning figurative language, and specifically that related to Satire. Section 3 presents the corpus obtained from Twitter. The overall design of the proposed approach is described in Section 4. In Section 5, we describe the experiments conducted to evaluate our method and present the results obtained. In Section 6, we discuss the performance of our method and compare it with those found in related works. Finally, our conclusions and future work are presented in Section 7.

2. Related Work

The literature related to this field considers the automatic detection of figurative language to be a difficult problem, and it has been addressed in only a few studies. With regard to sarcasm and irony, Poria et al.[4] proposed models based on a pre-trained convolutional neural network in order to extract sentiment, emotion and personality features for the detection of sarcasm. The authors carried out their experiments using three datasets, two of which were obtained from the work presented in Ptáček et al. [16], and the other of which was obtained from Sarcasm Detector. Their experimental results showed that, in the case of sarcasm detection, the baseline features outperformed the pre-trained features. However, the combination of pre-trained features and baseline features outperformed both of them when considered separately. Davidov, Tsur& Rappoport [5] used a semi-supervised sarcasm identification algorithm (SASI). This algorithm employed two modules: 1) semi-supervised pattern acquisition to identify sarcastic patterns that serve as features for a classifier, and 2) a classification stage that classifies each sentence into a sarcastic class. Moreover, the experiments were performed on two datasets: tweets collected from Twitter and Amazon product reviews. González-Ibáñez, Muresa & Wacholder [6], meanwhile, presented an empirical study on the use of lexical and pragmatic factors as a means to distinguish sarcasm from positive and negative sentiments expressed in Twitter messages. The authors developed a corpus that included only sarcastic utterances that had been explicitly identified as such by the composer of the message. They provided a report concerning the difficulty of distinguishing between sarcastic tweets and tweets that are straightforwardly positive or negative. The results showed the difficulty experienced by both humans and machine learning

¹<http://www.internetworldstats.com/stats7.htm>

methods when attempting to classify sarcasm. Filatova[7] described an experiment concerning corpus generation whose goal was to obtain Amazon product review pairs that could be used to analyze the notion of sarcasm in texts and to train a sarcasm detection system. Justo et al. [8] proposed various methods whose objective was to discover an appropriate set of features for different types of social language. They tested a range of different feature sets by using two different classifiers. The results showed that the sarcasm detection task benefitted from the inclusion of linguistic and semantic information sources, while nasty language was more easily detected by using only a set of surface patterns or indicators. Sulis et al. [9] investigated the use of figurative language in Twitter. Messages explicitly tagged by users as #irony, #sarcasm and #not were analyzed in order to test the hypothesis as regards dealing with different linguistic phenomena. They took into account emotional and affective lexical resources, in addition to structural features, with the aim of exploring the relationship between figurativity, sentiment and emotions at a finer level of granularity. Bouazizi & Ohtsuki [17] proposed an efficient means to detect sarcastic tweets, and studied how to use this information (i.e., whether or not the tweet is sarcastic) so as to enhance the accuracy of sentiment analysis. The main purposes for which sarcasm is used in social networks were also identified. Finally, they studied the added value of the different sets of features used in terms of precision. Delmonte & Stingo [18] presented a computational work whose intention was to detect satire/sarcasm in long commentaries on Italian politics. They used the lexical features extracted from the manual annotation based on Appraisal Theory from some 30,000 word texts. A manual annotation phase carried out on 112 texts by two well-known Italian journalists was presented. After a first experimentation phase based on the lexical features extracted from the XML (eXtensible Markup Language) output files, they proceeded to retag lexical entries by dividing them into two subclasses: figurative and literal meaning. Tsonkov & Koychev [19] described and proposed three different ways in which to automatically detect double meaning in English texts and proposed 9 heuristic rules for this purpose. Five extra features were then employed, such as the number of words containing positive sentiment and negative sentiment, Intensity score, the number of adjectives and the length of the opinion, in order to improve the accuracy of the rule-based approach. Karoui, Zitoun & Moriceau [20] proposed a model with which to identify irony in implicit oppositions in the French language. Sequential Minimal Optimization (SMO) was used, along with the Weka toolkit with standard parameters. Other learning algorithms (naive bayes, decision trees, logistic regression) were also evaluated, but the results were not as good as those obtained with SMO.

Satire detection has often been studied in literature, but a computational approach has not usually been used [13, 21, 22, 23, 24,25]. For instance, Ahmad et al. [21] proposed a method for satire detection, which is based on the use of machine learning techniques, and used a bag-of-words along with feature weighing in order to train a SVM (Support Vector Machine) algorithm. The experiments were carried out on a news corpus. The newswire documents were randomly sampled from the English Gigaword Corpus and the satire data set was selected from The Orion. Burfoot & Baldwin [22] presented an approach that could be used to determine whether a newswire article is “true” or satirical. They based their experiments on the use of SVM algorithms by considering several features such as bag-of-words, lexical features, and semantic validity. The results showed that the combination of SVMs with BNS (Bi-Normal Separation) feature scaling achieved high precision and lower recall, and that the inclusion of the notion of validity achieved the best overall F-score. Barbieri, Ronzano & Saggion [23], meanwhile, presented a system for the automatic detection of Spanish satirical news. The system developed relies on linguistic features to classify tweets. The experiments demonstrated that the model outperforms a word-based baseline as regards detecting whether or not a tweet is satirical. Owais, Nafis & Khanna [24] proposed an approach for use in the classification of online news articles as satirical or true by using the SVM (Support Vector Machine) classification method and bag-of-words features. They additionally supplemented the bag-of-words model with feature weighting by using two methods: 1) Binary feature weights, and 2) Bi-normal separation feature scaling. The corpus consisted of a total of 2500 true news documents and 110 satirical news articles. Rubin et al. [13] proposed an SVM-based algorithm, enriched with 5 predictive features: Absurdity, Humor, Grammar, Negative Affect, and Punctuation. A combination of these was tested on 360 news articles. The dataset was collected in 2 sets. The first set was collected from 2 satirical news sites (The Onion and The Beaverton) while the second was taken from 2 legitimate news sources (The Toronto Star and The New York Times). Barberi, Ronzano & Saggion [25] proposed an approach for the

detection of news satire in Twitter in English, Spanish, and Italian. A set of features describing lexical, semantic and usage-related properties of the words in each Tweet was used for this purpose.

We should state that our work differs from the existing literature for two main reasons: (i) The aforementioned proposals are based on different feature extraction approaches: some are based on a bag-of-words model while others explore more sophisticated features, such as linguistic and affective features or punctuation marks. However, these features have been selected on the basis of other studies, and on the assumption that the features have a great influence on sarcasm or satire detection. We have, therefore, analyzed a broader set of features related to psycholinguistic aspects, punctuations marks, personal concerns and spoken categories in order to discover important features that may help determinate whether or not a tweet contains satire, and (ii) We have carried out an analysis of a set of satirical tweets from Mexico and another from Spain in order to determine whether there is a significant difference between the patterns of features used to detect satire in these two cultures.

3. Corpus

This study has been carried out using a dataset concerning satirical and non-satirical news from Twitter accounts. Since satire detection has appeared only recently, few Spanish language datasets were available. Nonetheless, a corpus for Spanish satirical detection is presented in Barberi et al. [23]. However, the corpus provides only information regarding the Spanish of Spain. We have therefore decided to generate our own corpus, since our aim is also to carry out an analysis of Mexican satire. The tweet collection from the Twitter Streaming API was built using Twitter4J², a Java-based library that facilitates the usage of Twitter API. Tweets were retrieved from eight Twitter accounts, four of which are satirical (two from Spain and two from Mexico) and four of which are non-satirical (two from Spain and two from Mexico) (see Table 1). We obtained a total of 10,000 satirical tweets (5000 from Spain and 5000 from Mexico), and 10,000 non-satirical tweets (5000 from Spain and 5000 from Mexico).

Table 1. Twitter accounts regarding satirical and non-satirical news

	Country	Account
Satirical	Mexico	El deforma (@eldeforma) El dizque (@eldizque)
	Spain	El jueves (@eljueves) El mundo today (@elmundotoday)
Non-satirical	Mexico	El universal (@El_Universal_Mx) Excelsior (@Excelsior)
	Spain	El país (@el_pais) El mundo (@elmundo)

Once the tweets had been collected, we applied automatic filtering to remove retweets, duplicates, tweets written in non-Spanish languages, and tweets with only URLs. Finally, we performed a manual review of the filtered tweets in order to ensure that those obtained were relevant to our study. This resulted in 6829 satirical tweets (3107 from Mexico and 3722 from Spain) and 7716 non-satirical tweets (4561 from Mexico and 3155 from Spain). Our final corpus contains a total of 5000 satirical tweets and 5000 non-satirical tweets owing to the fact that we selected only 2500 satirical tweets from Spain, 2500 satirical tweets from Mexico, 2500 non-satirical tweets from Spain, and 2500 non-satirical from Mexico in order to obtain a balanced corpus.

Figure 1 shows an example of a non-satirical tweet, while Figure 2 shows an example of a satirical tweet from our corpus. We share this dataset as a list of Tweet IDs since, according to the Twitter privacy policy, the content of tweets cannot be shared.³

²<http://twitter4j.org/>

³<http://dis.um.es/~valencia/SatiricalTwitterDataSet.zip>

Figure 1 non-satirical tweet

Figure 2 satirical tweet

4. Our approach

The satire detection method proposed is divided into three main steps: (1) Text pre-processing, (2) Feature extraction, and (3) Training of the machine learning algorithm. The approach presented was tested on the Spanish language by using a corpus of tweets concerning satirical and non-satirical news from Mexican and Spanish Twitter accounts. Figure 3 shows the flow diagram of our satirical detection approach. This first step involves pre-processing the text in the corpus in order to clean and correct it. The second step consists of extracting the psychological and linguistic features by means of LIWC. The last step consists of training the machine learning classification algorithms, such as SMO, BayesNet and J48.

Figure 3 Flow diagram of our satire detection approach

4.1 Text pre-processing

The language used in Twitter is mainly characterized by an informal and free writing style [26], in which initialisms, shortenings, homophonic confusion, character repetition and the misuse of uppercase letters are commonly used to either save characters or denote emphasis in tweets. Twitter users frequently make spelling and grammar mistakes and create short texts that are difficult to analyze. We dealt with these phenomena of lexical variation by using the work presented in [27] as a basis.

The first step of our approach consists of a tokenization process of all tweets by using the Twokenize tool [28], a Twitter-specific tokenizer that not only divides the text into tokens but also allows the detection of entities of interest for this approach, namely mentions, hashtags, and URLs.

Any items that cannot assist during the subsequent phase are then removed with the objective of reducing the noise in the dataset. The following tasks are performed for each tweet:

1. Delete mentions and replies to other users' tweets, which are represented by means of strings starting with @.
2. Remove URLs, i.e., strings starting with http://
3. The “#” character is removed from all hashtags because often, only the remainder of the string forms a legible word that contributes to a better understanding of the tweet [29].

The use of abbreviations and shorthand notations is common in tweets, given the limitation of 140 characters. In this respect, the second step consists of replacing elements with their expansions. For instance, "también" (also) instead of "tb", "por favor" (please) instead of "xfa". This task is performed using the NetLingo dictionary[30].

Finally, the spelling errors are corrected. This is done using Hunspell [31], an open source spell checker and morphological analyzer based on MySpell and designed for languages with a rich morphology[32]. For instance, "palabar", "daibetes", and "compelto" are corrected to "palabra" (word), "diabetes" (diabetes), and "completo" (complete, full), respectively.

As mentioned above, the base resources used for this task are Hunspell, Twokenize, and NetLingo. These resources were selected because they have been widely and successfully used to pre-process tweets in other research works, such as[26],[27],[33],[34], [35], [36], [37].

4.2 Feature extraction

We used LIWC (Linguistic Inquiry and Word Count) in order to obtain the psycholinguistic features of the texts and explore which LIWC categories can provide patterns for satire detection. LIWC categories have been used in sarcasm classification [8], satire classification [13][14] and similar tasks, such as identifying humorous texts [12]. However, in the case of satire detection only features such as negative emotions, negations, quantifiers, certainty, tentative, exclusion, inclusion, discrepancy, causation, past, present and punctuation marks have been investigated. The authors of the aforementioned works selected a group of features based on previous studies in literature and by assuming that the features would be discriminating as regards satirical detection. In this respect, we have examined the complete list of the 72 categories available in LIWC in order to detect which of the features for figurative language mentioned in literature and other LIWC features may contribute to satire detection.

LIWC was specifically developed to provide an efficient method with which to study psycholinguistic concerns, and has been considerably improved since its first version. An updated revision of the original version was presented in Pennebaker et al. [38], namely LIWC2015. LIWC contains dictionaries in several languages such as English, Dutch, Portuguese, Arabic and Spanish, to mention but a few. These dictionaries have been translated from the English version. In this work, we used the 2007 Spanish version [39], which was overseen by a native speaker of Mexican Spanish who worked closely with a Colombian Spanish speaker. The final version involved the collaboration of a native Spanish speaker from Spain. This dictionary is composed of 7515 words and word stems. Each word can be classified into one or more of the 72 categories included by default in LIWC. These categories are additionally classified into five main sets: (1) linguistic processes; (2) psychological processes; (3) personal concerns; (4) spoken categories; and 5) Punctuation. Table 2 shows some examples of the LIWC categories. The full list of categories is presented in Ramírez Esparza et al. [39].

Table 2. LIWC categories

Set	Categories	Examples
Linguistic process (LP)	totpron/ total of pronouns	Yo, nosotros, tu
	artículo/Articles	Él, la, los, las
	Negacio/negations	No, nunca
Psychological process (PP)	Emopos	Feliz, bonito, bueno
	Emoneg	Odio, enemigo, feo
	Meccog/cognitive process	Causa, saber, debería
Personal concerns (PC)	Trabajo/work	Empleado, jefe, Carrera
	Dinero/money	Cambio, dinero, ganancia
	Relig/Religion	Dios, cielo, iglesia
Spoken categories (SC)	Asentir/assent	Acuerdo, afirmación, genial
	Nofluen/nonfluencies	Ah, aja, eh
	Relleno/filler	Osea, uf
Punctuation marks (PM)	Periods, commas, colons, semicolons, question marks, exclamation marks, dashes, quotation marks, apostrophes, parentheses, other punctuation.	

As can be seen in Table 2, the first set consists of the linguistic process, which involves grammatical information such as the total of pronouns, articles, negations, word counts and auxiliary verbs, among others. The second set contains the psychological process, which is able to estimate positive emotions, negative emotions, social processes and cognitive processes, among others. Within this dimension, the emotion or affective processes use subdictionaries that gather words selected from several sources such as the PANAS [40] and Roget's Thesaurus, which are subsequently rated by groups of three judges working independently. The third set consists of word categories related to personal concerns intrinsic to the human condition. The fourth set contains "spoken categories", and has been included in order to accommodate certain dimensions of spoken language. Although LIWC has not been

designed for the spoken language, the authors have found it useful when analyzing conversations and interviews. Finally, the fifth set consists of twelve punctuation categories (periods, commas, colons, etc.). With regard to parentheses, LIWC counts pairs of parentheses. Moreover, the “Other Punctuation” category contains all those punctuation marks not included in the other punctuation categories, along with ASCII characters from 33-47, 58-64, 91-96 and 123-126.

In order to carry out the text processing with LIWC, we converted the corpus into plain text files, after which each tweet was analyzed individually. LIWC carries out the text processing in three steps. First, the LIWC text analysis module compares each word in the text with a user-defined dictionary (in this case, we selected the Spanish dictionary). For instance, let us consider the tweet presented in Figure 4, “TV Azteca lanza “Plim”, la plataforma digital donde podrás volver a ver tus programas favoritos de la televisora - TV Azteca launches "Plim" a digital platform on which you can re-watch your favorite television programs". In this step, LIWC found fifteen of the eighteen words from the text (see Figure 5).

Figure 4. Tweet to analyze

TV Azteca lanza “Plim”, la plataforma digital donde podrás volver a ver tus programas favoritos de la televisora

Figure 5. Words found in the LIWC dictionary

The dictionary then identifies which words are associated with the LIWC categories. In addition to counting whole words, LIWC also counts word stems. Word stems are partial words that end with an asterisk (*) in the dictionary. The usage of an asterisk at the end of the word indicates that LIWC ignores all subsequent letters. Consequently, “favorit*” will include the words "favorito", "favorita", "favoritos", "favoritas". Figure 6 presents an extract of the results of the previous example. As can be seen, the words are categorized as follows: “lanza” as a cognitive process (“MecCog”), “la” as a pronoun (“TotPron”) and an article (“Articulo”), “ver” as a cognitive process (“MecCog”), “tus” as a pronoun (“TotPron”) and “favoritos” as a positive emotion (“EmoPos”). LIWC therefore finds that three pronouns (“la”, “tus”, and “la”), two articles (“la”, “la”), one positive emotion (“favoritos”), and two cognitive mechanisms (“lanza” and “ver”) are used.

	TotPron	Articulo	EmoPos	MecCog
TV				
Azteca				
lanza				x
plim				
la	x	x		
plataforma				
digital				
donde				
podrás				
volver				
a				
ver				x
tus	x			
programas				
favoritos			x	
de				
la	x	x		
televisora				

Figure 6. Categorized words

Once the processing module has read and accounted for all the words in a given text, the percentage of total words that match each of the dictionary categories is subsequently calculated. To continue with the excerpt from the previous example, LIWC analyzed a single tweet consisting of 18 words and found that three pronouns, two articles, one positive emotion, and two cognitive mechanism words are used, to mention but a few. These numbers are then converted into percentages (see Figure 7). The resulting feature vector therefore consists of the features with their respective percentages, with the exception of the summary variables WC (Total word count), WPS (Words per sentence), Sixltr (words longer than 6 letters), and DIC (Dictionary word count).

Filename	TotPron	Articulo	EmoPos	MecCog
706547424177889280.txt	16.67	11.11	5.56	11.11

Figure 7. Extract of the feature vector

4.3 Machine Learning

The final phase of the approach proposed consists of training the machine learning classification algorithms. In this work, we used WEKA [41], which is a collection of machine learning algorithms that can be used for data pre-processing, classification, regression, clustering, association rules and visualization.

Our work is based on machine learning classification algorithms known as classifiers. These classifiers allow the creation of models according to the data and purpose of analysis. The classifiers are categorized into seven groups: Bayesian (Naïve Bayes, Bayesian nets, etc.), functions (linear regression, SMO, logistic, etc.), lazy (IBk, LWL, etc.), meta-classifiers (Bagging, Vote, etc.), miscellaneous (SerializedClassifier and InputMappedClassifier), rules (DecisionTable, OneR, etc.) and trees (J48, RandomTree, etc.).

We have used Auto-WEKA to address the problem of learning algorithm selection and the tuning of its hyperparameters. In order to select the best classifier for this specific problem, Auto-WEKA carries out a combined algorithm selection and hyperparameter optimization on the regression and classification algorithms provided by WEKA. Given the Mexican and Spanish datasets, Auto-WEKA therefore explores hyperparameter settings for the algorithms in order to determine the best algorithm along with its corresponding parameter configuration. Auto-WEKA then recommends which method is, in this case, best for the detection of satirical tweets. Figure 8 presents the results obtained by Auto-WEKA for both datasets.

best classifier: weka.classifiers.functions.SMO
arguments: [-C, 1.0574050316054016, -N, 0, -K, weka.classifiers.functions.supportVector.NormalizedPolyKernel -E 4.819706475007974]

Figure 8. Results obtained by Auto-WEKA

where the C parameter represents a balance between the model complexity and the approximation error; N indicates the filter to apply to the training data, whose values can be: Normalize training data (0), Standardize training data (1) or No normalization/standardization (2); K indicates the kernel to use, in this case the NormalizedPolyKernel, and E determines the exponent value.

In order to validate whether the SMO classifier performs better than other algorithms employed in text classification problems[42][43][44][45], we performed the supervised training process by considering the Bayes Network learning algorithm (BayesNet), and the C4.5 decision tree (J48) classifiers. These algorithms were selected because they have been used in data classification problems[46][47][48] and have obtained encouraging results. Figure 9 depicts the machine learning workflow.

Figure 9. Machine learning workflow

The objective of the classifier training phase is to build a model based on the analysis of the instances. This model is represented by means of classification rules, decision trees or mathematical formulae. The models generated are used to classify unknown data, and our model is specifically able to distinguish between satirical and non-satirical texts. We performed a set of experiments with the aim of measuring the effectiveness of our approach for the detection of satirical tweets. These experiments are described in detail below.

5. Experimental evaluation

We have measured the performance of our method using well-known metrics: precision, recall, F-measure, accuracy and ROC Area. Recall is the proportion of actual positive cases that were correctly identified. This score is obtained using equation 1, where TP (True Positive) are those satirical tweets that the system classified as such and FN (False Negatives) are those satirical tweets that the system was not able to classify as such. Precision represents the proportion of positive cases identified that are real positives. The precision value is obtained by means of equation 2, where FP (False Positives) are the non-satirical tweets that the system classified as being satirical. The F-measure (see equation 3) is the harmonic mean of precision and recall, and is computed by multiplying the constant 2 by the product of precision and recall divided by their sum. Finally, accuracy is a weighted arithmetic mean of precision and inverse precision (weighted by Bias), in addition to being a weighted arithmetic mean of recall and inverse recall (weighted by prevalence) [49]. In this work, the accuracy score is obtained by using equation 4, where TP (True Positives) are those satirical tweets that the system classified as such, TN (True Negatives) are those non-satirical tweets that the system did not classify as being satirical, and Total population comprises both satirical and non-satirical tweets.

$$(1) \text{ Recall} = \frac{TP}{TP + FN}$$

$$(2) \text{ Precision} = \frac{TP}{TP + FP}$$

$$(3) F - measure = 2 * \frac{Precision * Recall}{Precision + Recall}$$

$$(4) Accuracy = \frac{TP + TN}{Total\ population}$$

LP	67.60	67.60	67.60	67.60	69.30	65.30	65.30	65.30	65.30	67.10
PP	63.10	63.00	62.90	63.00	64.30	61.90	61.90	61.80	61.90	63.20
PC	59.50	58.70	57.40	58.70	60.00	60.10	59.30	58.40	59.30	60.60
SC	56.70	52.30	37.70	52.30	52.30	64.00	53.10	39.00	53.10	52.90
PM	76.00	74.40	73.90	74.40	76.90	74.90	78.10	77.80	78.10	78.10

The second experiment led us to attain all the possible pair combinations obtained from the five sets. In this experiment, ten vectors were constructed (see sub-section 4.2), each of which considered only the features of the categories to be combined.

Table 4 shows that these results are better than those presented in Table 3. Note that it is not sufficient to use only one category of features. As can be observed in Table 4, the results are increased by simply adding the features from the linguistic and emotional categories. Conversely, the spoken categories and personal concerns are not as relevant as in the previous experiment. Furthermore, with regard to classifiers, SMO also obtained the best classification results in this experiment.

Table 4. Comparison of classification methods by using different feature sets

Satirical/Non-Satirical Mexico						Satirical/Non-Satirical Spain				
	P	R	F1	ACC	ROC	P	R	F1	ACC	ROC
SMO										
LP-PP	81.10	81.10	81.10	81.10	81.10	73.70	73.60	73.60	73.60	73.60
LP-PC	77.20	77.20	77.20	77.20	77.20	72.10	72.00	72.00	72.00	72.00
LP-SC	76.20	76.20	76.20	76.20	76.20	73.80	71.70	71.70	71.70	71.70
LP-PM	83.10	82.90	82.90	82.90	82.90	84.50	83.80	83.70	83.80	83.80
PP-PC	74.00	73.90	73.90	73.90	73.90	69.40	69.40	69.30	69.40	69.40
PP-SC	72.20	72.20	72.20	72.20	72.20	67.20	67.20	67.20	67.20	67.20
PP-PM	79.70	79.60	79.50	79.60	79.60	82.50	82.00	81.80	81.80	82.00
PC-SC	63.30	62.90	62.40	62.90	62.90	62.70	62.50	62.20	62.50	62.50
PC-PM	78.00	77.40	77.20	77.40	77.40	82.60	81.30	81.00	81.30	81.30
SC-PM	78.20	77.00	76.60	77.00	77.00	82.10	80.50	80.10	80.50	80.50
BayesNet										
LP-PP	69.10	69.10	69.10	69.10	75.00	66.00	66.00	66.00	66.00	71.30
LP-PC	67.80	67.80	67.80	67.80	73.70	65.90	65.90	65.90	65.90	71.10
LP-SC	67.20	67.20	67.20	67.20	72.70	65.80	65.80	65.80	65.80	70.80
LP-PM	74.70	74.50	74.50	74.50	82.80	74.70	74.50	74.50	74.50	82.80
PP-PC	65.50	65.40	65.40	65.40	70.30	63.00	62.90	62.90	62.90	67.10
PP-SC	64.00	64.00	64.00	64.00	68.30	62.20	62.20	62.20	62.20	66.50
PP-PM	74.40	74.20	74.20	74.20	80.80	75.40	75.20	75.10	75.20	82.70
PC-SC	58.60	57.70	56.10	57.70	59.70	61.40	57.90	53.30	57.90	60.10
PC-PM	73.50	72.80	72.50	72.80	78.70	78.20	76.70	76.30	76.70	83.60
SC-PM	74.90	73.80	73.40	73.80	76.20	78.90	77.70	77.40	77.70	83.20
J48										
LP-PP	68.80	68.80	68.80	68.80	69.70	66.80	66.80	66.80	66.80	66.90
LP-PC	66.70	66.70	66.70	66.70	68.10	65.70	65.70	65.70	65.70	65.90
LP-SC	67.30	67.30	67.30	67.30	68.90	66.00	66.00	66.00	66.00	67.70
LP-PM	78.30	78.20	78.20	78.20	79.10	78.30	78.20	78.20	78.20	79.10
PP-PC	66.00	65.90	65.90	65.90	67.30	62.30	62.20	62.20	62.20	62.60
PP-SC	64.10	64.00	63.80	64.00	65.50	62.30	62.10	62.00	62.10	64.20
PP-PM	75.70	75.60	75.60	75.60	79.80	76.80	76.70	76.70	76.70	77.90
PC-SC	60.30	59.40	58.10	59.40	60.70	60.80	60.20	59.50	60.20	61.90
PC-PM	75.10	74.40	74.10	74.40	78.70	78.00	77.30	77.10	77.30	81.80
SC-PM	76.00	74.40	73.90	74.40	76.90	79.90	78.60	78.30	78.60	81.40

The third experiment included all the feature sets. The vector that was constructed for this purpose therefore had a dimension of 72 (see sub-section 4.2), which included all the features.

Table 5 shows that SMO obtained the best results with a precision of 85.5%, a recall of 85.5%, an F-measure of 85.5% and an accuracy of 85.5% for Mexico, and a precision of 84.6%, a recall of 84.0%, an F-measure of 84.0% and an accuracy of 84.0% for Spain.

Table 5. Comparison of classification methods by using all feature sets

Satirical/Non-Satirical Mexico						Satirical/Non-Satirical Spain				
	P	R	F1	ACC	ROC	P	R	F1	ACC	ROC
SMO										
ALL	85.50	85.50	85.50	85.50	85.50	84.60	84.00	84.00	84.00	84.00
BayesNet										
ALL	75.70	75.60	75.60	75.60	75.60	73.40	73.40	73.40	73.40	81.10
J48										
ALL	75.20	75.20	75.20	75.20	76.10	77.40	77.40	77.40	77.40	76.70

The results obtained in Table 3 show that the most relevant categories are those regarding punctuation marks (PM) and linguistic process (LP). The results shown in Table 4 also show that the combination of the features of the relevant categories provides better results. For instance, if only the punctuation marks (PM) and the SMO classifier are used, then F-measures of 76.60% and 80.30% are obtained for Mexico and Spain, respectively. However, if LP and PM are combined, then the F-measure is increased to 82.90% and 83.70% for Mexico and Spain. Finally, Table 5 indicates that the experiments with all categories provided better results, which indicates that some features contained in the other categories are also relevant, such as those contained in the psychological process (PP).

Finally, we used a feature selection stage in order to understand the contribution that each individual feature makes to each model, and to obtain the most discriminant features for satire detection. This was done by ranking all available features according to the information gain criterion. Figure 10 shows the 13 most relevant features as regards detecting satire for Spain and Mexico, namely adverbs (“Adverb”), affect (“Afect”), apostrophe (“Apostro”), certainty (“Certeza”), quantifiers (“Cuantif”), colons (“Colon”), positive words (“EmoPos”), exclamation marks (“Exclam”), informal (“Informal”), cognitive mechanisms (“MecCog”), present tense (“Present”), quotation marks (“Quote”), and social (“Social”). The best ranked features are related to the linguistic process, the psychological process and punctuation marks. This confirms the relevance of this kind of features for satire detection.

Figure 10. Information Gain values for the 13 best ranked features

6. Discussion of Results

The results show that our approach provides encouraging results as regards the detection of satirical

and non-satirical tweets. They also confirm that linguistic features, emotional features and punctuation marks are important when attempting to detect satire.

As can be observed in Figure 11, with regard to the linguistic features, the results demonstrate that satirical tweets contain more adverbs, quantifiers and present tenses than non-satirical tweets. In the case of the psychological process, five features were significant predictors: social process, affective process, positive emotions, cognitive process and certainty. Finally, with regard to punctuation marks, colons, exclamation marks, apostrophes and quotation marks were significant predictors. The conclusions that can be drawn from these results are:

Figure 11. Mean percentage of words for satirical and non-satirical tweets

- Satirical tweets contain more positive words (“EmoPos”) in order to change the meaning of a negative statement, i.e., a social problem, abuses, or any outrageous situation that the user wishes to disclose. For instance, in the tweet presented in Figure 2 “Confirman que mensaje de Roger Waters llegó al corazón de Peña. México ahora es un mejor país. Se acabó la violencia y hay opulencia- It is confirmed that Roger Water’s message reached Peña’s heart. Mexico is now a better country. There is no more violence and opulence flourishes” is composed of positive words. However, this tweet is in reality attempting to show indignation at the situation that the country is confronting thanks to bad government.
- Cognitive mechanisms (“MecCog”) (e.g., cause, know, ought) are indicative of more complex language [52]. We therefore attribute the high rate of the cognitive process to the fact that satire is more complex and difficult to understand than literal language. For instance, the tweets presented in Figure 12 contain several words, such as “sabe”, “qué”, “sin”, “pero”, “prefiere”, “preguntar”, “igual”, “porque”, “que”, “siempre”, and “pensáis”, which belong to the cognitive mechanism.

Figure 12. Satirical tweets containing words related to cognitive mechanism

- We found that four categories: certainty ("Certeza"), informal ("Informal"), affect ("Afect") and quantifiers ("Cuantif") are strongly related to hyperbole, which is a recurring phenomenon in satirical utterances in which a real situation is exaggerated until it becomes absurd. We ascribe this to the following: 1) Certainty words such as "always" and "never" are some examples of words that generally exaggerate a situation; 2) Informal language is more narrative, emotional and interactive. In literature, these features have been associated with exaggeration [53][54]; 3) The findings in McCarthy & Carter's work [54] revealed that exaggeration is associated with a set of affective features, and 4) Mass quantifiers such as masses, loads and tons are features that usually appear in hyperbolic contexts. The tweets presented in Figure 13 are examples of hyperbole in which a situation is exaggerated. In the first example, informal language and certainty words such as "siempre" were used with this objective. In the second tweet, affective words such as "teme" and quantifiers such as "todo" were used to increase the veracity.

Figure 13. Examples of tweets containing hyperbole

- Adverbs ("Adverb") are also a good indicator of satirical tweets, and are used to exaggerate or to minimize a statement [55]. In the case of the tweets presented in Figure 14, the adverbs "tan" and "muy" were used to intensify the meaning.

Figure 14. Satirical tweets containing adverbs

- Punctuation marks, such as exclamation marks (“Exclam”), colons (“Colon”) and quotation marks (“Quote”), are frequently used to denote satire. Based on the findings of Sulis et al. [56] and Kovaz et al.[9], we conclude that satire shares these features with irony and sarcasm. The tweets presented in Figure 15 are clear examples of satirical tweets that contain punctuation marks such as exclamation marks (!), colons (:), and quotation marks (“”, «», «»).

Figure 15. Satirical tweets containing punctuation marks

Furthermore, the results show that the apostrophe (“Apostro”) is more commonly used in non-satirical tweets. For instance, Figure 16 shows two examples of non-satirical tweets in which apostrophes were used.

Figure 16. Non-satirical tweets containing punctuation marks

- The “Social” category is also relevant in these results. However, it is more content-dependent, i.e., news satire usually represents current social issues, and we do not therefore believe that it is truly discriminant as regards satire detection. For instance, the tweets presented in Figures 12, 13 and 15 are commonly related to social concerns such as politics.
- The situation of the “Present” category is similar to that of the “Social” category because it is context-dependent. For example, Skalicky & Crossley [14] found that the past tense was more discriminant in Amazon reviews owing to the fact that satirical texts presented a story or narrative related to the product being reviewed, and as such, more past tense markers were found. However, our results show that satirical news has more present tense markers. We attribute this to the fact that news frequently expresses current situations. For example, let us consider the tweets presented in Figures 12, 14 and 15. In these tweets, the writer satirizes current news (at the time of publication). As can be seen in Table 6, several verbs were identified as being in the present tense, which demonstrates that satirical news tweets are commonly written in this tense.

Table 6. Examples of satirical tweets written in the present tense

Tweet	Present category
“Rajoy no sabe por qué lleva un año sin trabajar pero prefiere no preguntar nada para no romper la racha”	Sabe, lleva, prefiere
“Igual el gobierno está vaciando la hucha de las pensiones porque debajo del dinero hay un Pokemon. Es que, joder, siempre pensáis lo peor.”	Está, hay, es, pensáis
“Sony lanza un teléfono muy resistente porque cree que eres imbécil y se te va a caer.”	Lanza, cree, eres, va
“Los anuncios de la Lotería son tan falsos que ya ni siquiera ponen personas de verdad en ellos”	Son, ponen
“Si escuchas el discurso de Pedro Sánchez al revés se oye el mensaje: “¡Ayuda, Albert	Oye, es

Rivera es un reptiliano!"

As can be observed in Figure 17, there are no significant differences between Mexican and Spanish satire. Although the same discriminant features were obtained by our approach for both datasets, there are differences with regard to the percentage used in the news satire. For instance, there is a great difference between Spanish and Mexican news satire with regard to the use of exclamation marks and cognitive mechanisms, since they are more frequently used in Mexican satire, i.e., Mexican writers commonly use exclamation marks to modify the tone of a statement and to denote satire in the text. What is more, the use of words related to cognitive mechanisms ("MecCog") such as "admitir", "afecta", "sabe", "pero", "algún", "causa" are more frequently used in Mexican than in Spanish news satire.

With regard to other features, such as social, present, positive emotions ("EmoPos"), quote, colon, informal, quantifiers ("Cuantif"), certainty ("Certeza") and adverb, there is no significant difference in their use. We therefore conclude that positive emotions, punctuation marks such as colons and quotes, informal words, and quantifiers are used with the same frequency in Mexican and Spanish satirical tweets. In addition, writers attempt to express current social issues, such as those which are political (see Figure 15), and features such as present and social therefore play an important role. Finally, adverbs and certainty words ("Certeza") such as "siempre", "jamás", "muy", "nunca" are used by both Spanish and Mexican writers to exaggerate or minimize a situation.

Figure 17. Mean percentage of words for Spain and Mexico

Figure 18 depicts the evaluation results obtained after running the three algorithms and employing the ROC curve. This figure has three curves corresponding to each algorithm. The thinnest curve represents the SMO algorithm, the thickest curve represents the BayesNet classifier and the other represents the J48 algorithm. As can be observed, the SMO curve is closer to the top-left point in both cases, i.e., in the case of both the Spanish and the Mexican datasets. This indicates that the SMO is more accurate than the other classifiers. The J48 and BayesNet curves are, meanwhile, closer to the diagonal, which indicates that these classifiers are less accurate. The results obtained match those reported by other works in literature, which have concluded that SVM is one of the best machine learning classification algorithms and outperforms algorithms such as J48, BayesNet, and MaxEnt, among others [42][43][44][45]. Furthermore, these evaluation results are justified by the analysis of several classifiers reported in [57], in which it has been clearly demonstrated that SVM models are more robust and accurate than other algorithms, including those compared in this work. Other facts that justify these results correspond to the robustness of SMO in high dimensional spaces and in sparse sets of samples, and most text categorization problems are linearly separable [58]. Unlike other classifiers, such as decision trees or logistic regressions, SVM assumes no linearity, and it can be difficult to interpret its results outside its accuracy values [59].

Figure 18 ROC curves for Mexican and Spanish satire

6.1 Comparison with related work

Table 6 shows the results obtained from similar proposals focused on satire detection that have been evaluated in terms of the F-measure. In general, the proposals based on the English language obtained better results (i.e., [21], [24] and [13]). We ascribe this to two major factors. First, English has a lower level of grammatical complexity than Spanish. In our work, as has been stated in Section 5, both punctuation marks and linguistic process sets are the most discriminant features. Although the punctuation marks make a significant contribution to the classification, other language-dependent linguistic features, such as “Adverb” or “Cuantif”, have a considerable influence on the results. Second, the NLP tools oriented toward English are more mature, extensive and available than those in other languages such as Spanish. With regard to the Spanish language, the results are similar, although our proposal outperformed the results of [23] and [25]. However, it is difficult to establish whether a work is better or worse than our proposal for two principal reasons: 1) the complexity of the domain, and 2) the language used.

Table 7. Comparison with related works

Proposal	Language	F-measure
Ahmad et al. [21]	English	84.0
Burfoot & Baldwin [22]	English	79.8
Barbieri, Ronzano & Saggion [23]	Spanish	81.4
Owais, Nafis & Khanna [24]	English	83.0
Rubin et al. [13]	English	87.0
Barbieri, Ronzano & Saggion [25]	English	76.3
	Spanish	81.6
	Italian	80.0
Our approach	Spanish	84.0
		85.5

We are therefore of the opinion that comparing the different approaches described in literature is difficult because none of the software applications are available. Indeed, the corpora used for each experiment differ significantly as regards content, size, topics and language. A fair comparison of two methods would require the usage of the same testing corpus, and it was for this reason that we carried out an exhaustive search for a standard corpus concerning satire in the Spanish language. This search was performed at well-known conferences and workshops such as SemEval, CLEF and WASSA. Despite our efforts, we found only datasets concerning irony and sarcasm in the English language. Although we found proposals focused on satire detection, such as [13, 21, 22, 23, 24, 25], it is not possible to make a fair comparison with them for two main reasons: 1) the proposals presented in [13, 21, 22, 24] are focused on another language, and 2) the corpora used for these experiments are not publicly available.

7. Conclusion and Future Work

The contribution of this work is twofold. First, we have collected a corpus composed of Spanish and Mexican satirical and non-satirical news from Twitter accounts. Second, we have provided a new satire detection method based on psycholinguistic features. We have also presented experiments whose objective was to evaluate the proposed method. Our proposal obtained encouraging results, with a Precision of 85.5%, a Recall of 85.5%, an F-measure of 85.5%, an Accuracy of 85.50% and a ROC area of 85.5% in the case of news from Mexico, and a Precision of 84.6%, a Recall of 84.0%, an F-measure of 84.0%, an Accuracy of 84.0% and a ROC area of 84.0% in the case of news from Spain. Furthermore, our results clearly confirm the usefulness of adopting the linguistic process, the psychological process and punctuation marks.

As future research, we have considered developing a new corpus in order to confirm the finding concerning “social” category, which we believe is content-dependent. We are also planning to explore the use of our approach with other languages such as English, Arabic, and French with the objective of determining whether the same pattern of features is useful to detect satire in other languages. Furthermore, we are interesting in carrying out a comparative analysis of sarcasm, irony, satire, with the aim of identifying the features that they share and determining whether they are similar or distinct phenomena. Finally, we have considered addressing idioms because they are frequently found in satirical text and we believe that they may contribute greatly to the detection of satirical text.

Acknowledgments

This work has been supported by the Spanish Ministry of Economy and Competitiveness and the European Commission (FEDER / ERDF) through project KBS4FIA (TIN2016-76323-R). María del Pilar Salas-Zárate and Mario Andrés Paredes-Valverde are supported by the National Council of Science and Technology (CONACYT), the Secretariat of Public Education (SEP) and the Mexican government. This work was also supported by Tecnológico Nacional de México (TecNM) and Secretariat of Public Education (SEP) through PRODEP (Programa para el Desarrollo Profesional Docente, in Spanish).

References

- [1] E. Cambria, B. Schuller, Y. Xia, and B. White, ‘New Avenues in Knowledge Bases for Natural Language Processing’, *Know-Based Syst.*, vol. 108, no. C, pp. 1–4, Sep. 2016.
- [2] E. Cambria, ‘Affective Computing and Sentiment Analysis’, *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, Mar. 2016.
- [3] C. Bosco, V. Patti, and A. Bolioli, ‘Developing Corpora for Sentiment Analysis: The Case of Irony and senti-TUT’, in *Proceedings of the 24th International Conference on Artificial Intelligence*, Buenos Aires, Argentina, 2015, pp. 4158–4162.
- [4] S. Poria, E. Cambria, D. Hazarika, and P. Vij, ‘A Deeper Look into Sarcastic Tweets Using Deep Convolutional Neural Networks’, in *Proceedings of COLING 2016*, 2016, pp. 1601–1612.
- [5] D. Davidov, O. Tsur, and A. Rappoport, ‘Semi-supervised Recognition of Sarcastic Sentences in Twitter and Amazon’, in *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, Stroudsburg, PA, USA, 2010, pp. 107–116.
- [6] R. González-Ibáñez, S. Muresan, and N. Wacholder, ‘Identifying Sarcasm in Twitter: A Closer Look’, in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, Stroudsburg, PA, USA, 2011, pp. 581–586.
- [7] E. Filatova, ‘Irony and sarcasm: Corpus generation and analysis using crowdsourcing’, in *In Language Resources and Evaluation Conf., LREC2012*, 2012, pp. 392–398.
- [8] R. Justo, T. Corcoran, S. M. Lukin, M. Walker, and M. I. Torres, ‘Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web’, *Knowl.-Based Syst.*, vol. 69, pp. 124–133, Oct. 2014.
- [9] E. Sulis, D. Irazú Hernández Farías, P. Rosso, V. Patti, and G. Ruffo, ‘Figurative messages and affect in Twitter: Differences between #irony, #sarcasm and #not’, *Knowl.-Based Syst.*, vol. 108, pp. 132–143, Sep. 2016.

- [10] M. E. Francis and J. W. Pennebaker, 'LIWC: Linguistic Inquiry and Word Count', Dallas, TX: Southern Methodist University, 1993.
- [11] J. Brubaker, F. Kivran-Swaine, L. Taber, and G. Hayes, 'Grief-Stricken in a Crowd: The Language of Bereavement and Distress in Social Media', in *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.
- [12] R. Mihalcea and S. Pulman, 'Linguistic Ethnography: Identifying Dominant Word Classes in Text', in *Computational Linguistics and Intelligent Text Processing*, A. Gelbukh, Ed. Springer Berlin Heidelberg, 2009, pp. 594–602.
- [13] V. L. Rubin, N. J. Conroy, Y. Chen, and S. Cornwell, 'Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News.', presented at the Proceedings of the Workshop on Computational Approaches to Deception Detection at the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-CADD2016), San Diego, California, 2016, pp. 7–17.
- [14] S. Skalicky and S. Crossley, 'A statistical analysis of satirical Amazon.com product reviews', *Eur. J. Humour Res.*, vol. 2, no. 3, pp. 66–85, Apr. 2015.
- [15] J. D. Campbell and A. N. Katz, 'Are There Necessary Conditions for Inducing a Sense of Sarcastic Irony?', *ResearchGate*, vol. 49, no. 6, pp. 459–480, Aug. 2012.
- [16] T. Ptáček, I. Habernal, and J. Hong, 'Sarcasm detection on Czech and English Twitter', in *COLING (2014)*, 2014, pp. 213–223.
- [17] M. Bouazizi and T. O. Ohtsuki, 'A Pattern-Based Approach for Sarcasm Detection on Twitter', *IEEE Access*, vol. 4, pp. 5477–5488, 2016.
- [18] R. Delmonte and M. Stingo, 'Detecting Satire in Italian Political Commentaries', in *Computational Collective Intelligence*, N. T. Nguyen, L. Iliadis, Y. Manolopoulos, and B. Trawiński, Eds. Springer International Publishing, 2016, pp. 68–77.
- [19] T. V. Tsonkov and I. Koychev, 'Automatic Detection of Double Meaning in Texts from the Social Networks', presented at the Proceedings of the 2015 Balkan Conference on Informatics: : Advances in ICT, 2015, pp. 33–39.
- [20] J. Karoui, B. Farah, V. Moriceau, N. Aussenac-Gilles, and L. Hadrich-Belguith, 'Towards a Contextual Pragmatic Model to Detect Irony in Tweets', 2015, pp. 644–650.
- [21] T. Ahmad, H. Akhtar, A. Chopra, and M. W. Akhtar, 'Satire Detection from Web Documents Using Machine Learning Methods', in *Proceedings of the 2014 International Conference on Soft Computing and Machine Intelligence*, Washington, DC, USA, 2014, pp. 102–105.
- [22] C. Burfoot and T. Baldwin, 'Automatic Satire Detection: Are You Having a Laugh?', in *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, Stroudsburg, PA, USA, 2009, pp. 161–164.
- [23] F. Barbieri, F. Ronzano, and H. Saggion, 'Is this Tweet Satirical? A Computational Approach for Satire Detection in Spanish', *Proces. Leng. Nat.*, vol. 55, no. 0, pp. 135–142, Sep. 2015.
- [24] S. T. Owais, T. Nafis, and S. Khanna, 'An Improved Method for Detection of Satire from User-Generated Content', vol. 6, no. 3, pp. 2084–2088, 2015.
- [25] F. Barbieri, F. Ronzano, and H. Saggion, 'Do We Criticise (and Laugh) in the Same Way? Automatic Detection of Multi-lingual Satirical News in Twitter', in *Proceedings of the 24th International Conference on Artificial Intelligence*, Buenos Aires, Argentina, 2015, pp. 1215–1221.
- [26] J. M. Coteló, F. L. Cruz, J. A. Troyano, and F. J. Ortega, 'A modular approach for lexical normalization applied to Spanish tweets', *Expert Syst. Appl.*, vol. 42, no. 10, pp. 4743–4754, Jun. 2015.
- [27] J. Vilares, M. A. Alonso, and D. Vilares, 'Prototipado Rápido de un Sistema de Normalización de Tuits: Una Aproximación Léxica', presented at the Tweet-Norm@ SEPLN 2013, 2013, pp. 39–43.
- [28] O. Owoputi, B. O'Connor, C. Dyer, K. Gimpel, N. Schneider, and N. Smith, 'Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters', *Proc. Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol.*, pp. 380–391, Jun. 2013.
- [29] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, 'Ontology-based sentiment analysis of twitter posts', *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4065–4074, Aug. 2013.
- [30] E. Jansen, *NetLingo: The Largest List of Chat Acronyms and Text Shorthand*. NetLingo Inc., 2014.

- [31] L. Németh, 'Hunspell'. [Online]. Available: <https://hunspell.github.io/>. [Accessed: 15-Sep-2016].
- [32] L. Németh, V. Trón, P. Halácsy, A. Kornai, A. Rung, and I. Szakadát, 'Leveraging the open source ispell codebase for minority language analysis', in *First Steps in Language Documentation for Minority Languages*, 2004, pp. 56–59.
- [33] A. F. Anta, L. N. Chiroque, P. Morere, and A. Santos, 'Sentiment Analysis and Topic Detection of Spanish Tweets: A Comparative Study of NLP Techniques', *Proces. Leng. Nat.*, vol. 50, no. 0, pp. 45–52, Apr. 2013.
- [34] F. H. Khan, S. Bashir, and U. Qamar, 'TOM: Twitter opinion mining framework using hybrid classification scheme', *Decis. Support Syst.*, vol. 57, pp. 245–257, Jan. 2014.
- [35] H. Cordobés, A. F. Anta, L. F. Chiroque, F. P. García, T. Redondo, and A. Santos, 'Graph-based Techniques for Topic Classification of Tweets in Spanish.', *IJIMAI*, vol. 2, no. 5, pp. 32–38, 2014.
- [36] M. T. Wiley, C. Jin, V. Hristidis, and K. M. Esterling, 'Pharmaceutical drugs chatter on Online Social Networks', *J. Biomed. Inform.*, vol. 49, pp. 245–254, Jun. 2014.
- [37] F. H. Khan, U. Qamar, and M. Y. Javed, 'SentiView: A visual sentiment analysis framework', in *International Conference on Information Society (i-Society 2014)*, 2014, pp. 291–296.
- [38] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, 'The Development and Psychometric Properties of LIWC2015', Sep. 2015.
- [39] N. Ramírez-Esparza, J. W. Pennebaker, F. A. García, and R. Suriá Martínez, 'La psicología del uso de las palabras: un programa de computadora que analiza textos en español', *The psychology of word use: a computer program that analyzes texts in Spanish*, Jun. 2007.
- [40] D. Watson, L. A. Clark, and A. Tellegen, 'Development and validation of brief measures of positive and negative affect: the PANAS scales', *J. Pers. Soc. Psychol.*, vol. 54, no. 6, pp. 1063–1070, Jun. 1988.
- [41] R. R. Bouckaert *et al.*, 'WEKA—Experiences with a Java Open-Source Project', *J. Mach. Learn. Res.*, vol. 11, p. 2533–2541, Sep. 2010.
- [42] B. A. Shboul, M. Al-Ayyoub, and Y. Jararweh, 'Multi-way sentiment classification of Arabic reviews', in *2015 6th International Conference on Information and Communication Systems (ICICS)*, 2015, pp. 206–211.
- [43] M. Kaya, G. Fidan, and I. H. Toroslu, 'Transfer Learning Using Twitter Data for Improving Sentiment Classification of Turkish Political News', in *Information Sciences and Systems 2013*, E. Gelenbe and R. Lent, Eds. Springer International Publishing, 2013, pp. 139–148.
- [44] L. Yijing, G. Haixiang, L. Xiao, L. Yanan, and L. Jinling, 'Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data', *Knowl.-Based Syst.*, vol. 94, pp. 88–104, Feb. 2016.
- [45] P. Moradi and M. Rostami, 'Integration of graph clustering with ant colony optimization for feature selection', *Knowl.-Based Syst.*, vol. 84, pp. 144–161, Aug. 2015.
- [46] J. Nahar, T. Imam, K. S. Tickle, A. B. M. Shawkat Ali, and Y.-P. P. Chen, 'Computational intelligence for microarray data and biomedical image analysis for the early diagnosis of breast cancer', *Expert Syst. Appl.*, vol. 39, no. 16, pp. 12371–12377, Nov. 2012.
- [47] L. Chen, L. Qi, and F. Wang, 'Comparison of feature-level learning methods for mining online consumer reviews', *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9588–9601, Aug. 2012.
- [48] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Ureña-López, and A. R. Montejo-Ráez, 'Sentiment analysis in Twitter', *Nat. Lang. Eng.*, vol. 20, no. 1, pp. 1–28, Jan. 2014.
- [49] M. del P. Salas-Zárate, R. Valencia-García, A. Ruiz-Martínez, and R. Colomo-Palacios, 'Feature-based opinion mining in financial news: An ontology-driven approach', *J. Inf. Sci.*, p. 165551516645528, May 2016.
- [50] S. Gao, C.-H. Lee, and J. H. Lim, 'An ensemble classifier learning approach to ROC optimization', in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, vol. 2, pp. 679–682.
- [51] F. Gurcan and A. A. Birturk, 'A Hybrid Movie Recommender Using Dynamic Fuzzy Clustering', in *Information Sciences and Systems 2015*, O. H. Abdelrahman, E. Gelenbe, G. Gorbil, and R. Lent, Eds. Springer International Publishing, 2016, pp. 159–169.
- [52] Y. R. Tausczik and J. W. Pennebaker, 'The Psychological Meaning of Words: LIWC and

- Computerized Text Analysis Methods', *J. Lang. Soc. Psychol.*, vol. 29, no. 1, pp. 24–54, Mar. 2010.
- [53] J. Lindey, 'Literal versus exaggerated always and never', *IJCL*, vol. 21, no. 2, pp. 219–249, 2016.
- [54] M. McCarthy and R. Carter, "'There's millions of them': hyperbole in everyday conversation", *J. Pragmat.*, vol. 36, no. 2, pp. 149–184, Feb. 2004.
- [55] D. Kovaz, R. J. Kreuz, and M. A. Riordan, 'Distinguishing Sarcasm From Literal Language: Evidence From Books and Blogging', *Discourse Process.*, vol. 50, no. 8, pp. 598–615, Nov. 2013.
- [56] R. J. Kreuz and G. M. Caucci, 'Lexical Influences on the Perception of Sarcasm', in *Proceedings of the Workshop on Computational Approaches to Figurative Language*, Stroudsburg, PA, USA, 2007, pp. 1–4.
- [57] H. Bhavsar and A. Ganatra, 'A Comparative Study of Training Algorithms for Supervised Machine Learning', *International Journal of Soft Computing and Engineering*, vol. 2, no. 4, pp. 74–81, 2012.
- [58] M. Rushdi Saleh, M. T. Martín-Valdivia, A. Montejó-Ráez, and L. A. Ureña-López, 'Experiments with SVM to classify opinions in different domains', *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14799–14804, Nov. 2011.
- [59] Y.-W. Chen and C.-J. Lin, 'Combining SVMs with Various Feature Selection Strategies', in *Feature Extraction*, I. Guyon, M. Nikraves, S. Gunn, and L. A. Zadeh, Eds. Springer Berlin Heidelberg, 2006, pp. 315–324.