Detecting Irony on Greek Political Tweets: A Text Mining Approach

Basilis Charalampakis, Dimitris Spathis, Elias Kouslis, and Katia Kermanidis
Department of Informatics, Ionian University
Corfu, Greece
{p11char, p11spat, p11kous, kerman}@ionio.gr

ABSTRACT

The present work describes the classification schema for irony detection in Greek political tweets. The proposed approach relies on limited labeled training data, and its performance on a larger unlabeled dataset is evaluated qualitatively (implicitly) via a correlation study between the irony that a party receives on Twitter, its respective actual election results during the Greek parliamentary elections of May 2012, and the difference between these results and the ones of the preceding elections of 2009. The machine learning results on the labeled dataset were highly encouraging and uncovered a trend whereby the volume of ironic tweets can predict the fluctuation from previous elections.

Keywords

Irony Detection, Text Mining, Twitter, Politics

1. INTRODUCTION

Irony is figurative linguistic element usually used to express an event with a semantic interpretation that is very different from its actual intended meaning. It is a challenging field for computational linguistics because of the high ambiguity and the difficulty to detect it. A language use is dynamic and creative, there is no consensual agreement on how to recognize an ironic statement, due to the high subjectivity involved.

In recent years, irony expression thrives on social media sites and particularly Twitter, because of the 140 characters restraint on the size of its postings, resembling the ever-existing one-liners. Being a public social medium, users realize that their thoughts might be read and reproduced by everyone, gaining popularity and followers. But this publicity, contrary to Facebook's real name policy, often has no direct consequences to their personal lives, since the majority of the users are writing anonymously, using an avatar.

This no-censorship state contributes to the freedom of expressing personal beliefs on tough, taboo, unpopular or controversial issues, part of which contains the political satire.

16th EANN workshops, September 25 - 28, 2015, Rhodes Island, Greece

http://dx.doi.org/10.1145/2797143.2797183 ...$15.00.

Political satire is a significant part of comedy which specializes in drawing entertainment from politics. Most of the times, it aims just to provide entertainment. By nature, it rarely offers a conducive view in itself; when it is used as part of criticism or jeer, it tends to simply establish the wrongness of things rather than provide tangible solutions.

The high topicality of Twitter, combined with the ephemerality of political news, forms a state which is described as 'echo chamber', a group-thinking effect on virtually enclosed spaces, amplified by repetition (Colleoni et al., 2014). As a result, the occasional user might write something political just to 'jump on the bandwagon', without an initial conscious aim to criticize. Studies focus on the simultaneous usage of Twitter and the TV on circumstances like a political debate, where meta-talk tweets reveal critical scrutiny of the agenda or 'the debate about the debate' (Kalsnes et al., 2014).

Considering the above, our empirical study tries to detect irony on a corpus of Greek political tweets by training a classifier, using appropriate linguistic features, some of which are proposed for the first time herein for irony detection. Our goal is to find a relation between the ironic tweets that refer to the political parties and leaders in Greece in the pre-election period of May 2012, and their actual election results. Unlike most previous attempts to irony detection, the proposed approach relies on limited labeled data, and its performance on a larger unlabeled dataset is evaluated implicitly by studying the difference in the election outcome for political parties and leaders in the Greek parliamentary elections of 2009 and 2012, and how this difference relates to the irony content of the tweets posted during the pre-election period of May 2012.

The remainder of this paper is organised as follows: In section 2, we present the related literature on the topics of irony detection, Twitter sentiment analysis and political expression. The next sections (3 & 4) are dedicated to data preprocessing and its representation schema through the set of linguistic features that affect irony detection. The Test and Results section (5) describes the training procedure, the evaluation of the algorithms' performance and their test procedure on a large unlabelled dataset. An overview of the study limitations and future research prospects are described in section 6. Finally, on section 7, we conclude with a summary of the empirical study.

2. RELATED WORK

The greater part of the literature on irony detection in computational linguistics is focused on English, but this is a first attempt to explore this area in the Greek language, to the authors' knowledge.

The most comprehensive study on the topic by Reyes et al. (2013) attempts to detect irony by examining the corpus on the following features: signatures, unexpectedness, style, and emotional scenarios. They used multiple datasets in order to evaluate their hypothesis and achieved a precision of 0.79 at best. Also a crisis management case study of the hashtag #Toyota is described.

A highly technical study by Rajadesingan et al. (2014) discovered an interesting aspect of Twitter usage, an 'orientation phase' in which the user is gradually introduced to irony as one gains followers. The threshold of this phase is one's 30 initial tweets.

The usual approach on similar irony detection studies on Twitter is to identify the two classes by hashtag analysis. However, this method creates noisy results with low accuracy (González-Ibáñez et al., 2011; Liebrecht et al., 2013).

Twitter lexical analysis on Greek tweets has been the main subject of the research by Kermanidis and Maragoudakis (2013), examining the sentimental tagging in a supervised environment. Their hypothesis is focused on the positive / negative distinction, using statistical techniques such as count and frequency distributions. The alignment between actual political results and web sentiment in both directions was investigated and confirmed. We use the same corpus of tweets in our study.

Apart from Twitter, similar techniques have been applied on Amazon reviews as well, making use of structured information of reviews versus the unstructured nature of Twitter. The accuracy results are encouraging due to the semi-supervised technique and the huge dataset Davidov et al. (2010).

Another study dealing with online opinion and reviews, again by Reyes et al. (2011), examined Amazon and Slashdot.com customer reviews trained on Naive Bayes, Decision Trees and Support Vector Machines. Accuracy results were satisfying and feature selection ranked as top the features of POS 3-gram (frequent sequence of trigrams) and Pleasantness (dictionary approach to pleasant and unpleasant words).

The novelty of our study lies on the leverage of these above techniques in Greek, based on the hypothesis that sarcasm and irony in Twitter messages may be linked to actual elections results. All text mining tools for extracting the learning features for the representation of the data have been developed from scratch, and are described in the following sections.

3. DATA PREPROCESSING

The dataset contains 61.427 Greek tweets collected on the week before and the week after the May 2012 parliament elections in Greece. The dataset is divided in 2 sub-datasets: parties and leaders. For each one, there are two sub-sections: before and after the elections. The dataset structure before the clean-up is presented in Tables 1 and 2 and is available[1] for research purposes.

[1]http://di.ionio.gr/hilab/doku.php?id=start:websent

Party	ANTARSYA	PASOK	DHMAR	KKE	ND	SYRIZA	XA	ANEL
Before Elections	1615	2909	1373	1588	792	5159	2731	791
After Elections	1628	9984	506	1594	1272	2294	3300	1109

Table 1: Number of tweets mentioning parties.

Leader	Tsipras	Kammenos	Samaras	Michaloliakos
Before Elections	3282	3511	4513	1059
After Elections	2992	1721	3171	2530

Table 2: Number of tweets mentioning leaders.

The first step was to eliminate the duplicate tweets in order to avoid frequency bias. The second step was to delete the useless, unstructured artifacts (unformated tweets) that were fetched with the Twitter API. In order to form a unibody test set, we merged the above sub-datasets. We decided to keep the tweets that contain links, because our hypothesis supports that tweets with links, for instance newspaper article tweets, are neutral, not ironic. After the cleanup, the tweets were 44.438.

The semantic analysis was assigned to Balkanet (Tufis et al., 2004), a greek edition of the WordNet (OMW: Open Multilingual Wordnet), a popular lexical database that groups words into sets of cognitive synonyms (synsets), each expressing a distinct concept.

Also, the Python natural language package, NLTK (Bird et al., 2009), was used in order to support Wordnet. The machine learning and training process was performed using the Weka software.[2]

4. FEATURES

We approached irony detection as a text classification problem. The decision if a tweet is ironic or not is a binary decision. We tag each tweet with five features, taking into consideration structural sentence formations and unexpectedness occurrences. Some of the features are designed to detect imbalance and unexpectedness, others to detect common patterns in the structure of the ironic tweets (like type of punctuation, length, emoticons). (Barbieri and Sagion, 2014) Our features are grouped into the following model:

- Spoken (spoken style applied in writings)
- Rarity (the frequency occurrences of the most rare words)
- Meanings (the number of Wordnet synsets as a measure of ambiguity)
- Lexical (punctuation, prosodic repeated letters, metaphors)
- Emoticons (smiley faces etc)

4.1 Spoken

The verbal irony in Twitter is often expressed as daily chats between virtual characters, using heavily dashes (-) and asterisks (*). Their occurrences in tweets count positively in our classifier. The use of spoken language is often

[2]http://www.cs.waikato.ac.nz/ml/weka

related to unexpectedness. In political context, dashes may be used to quote an actual quote, but the reason is usually ironic. The asterisk character showcases movements or non-verbal actions in tweets, such as *sigh*, adding an emotional level. If there is at least one of the above characters in the tweet, the value of the feature is 'true', otherwise it is 'false'. Thus, the spoken feature is binary.

4.2 Rarity

A frequency dictionary for all the words of the original dataset was created. The tweets are split into tokens, and token occurrences (excluding URL links) are counted.. Thus, we isolated the most rare words and limited the upper bound to three occurrences. The resulted frequency dictionary consists of 25.898 words. If a word had 1 occurrence, it was one of the most rare cases. If a word had 3 occurrences, it was less rare. In order to invert this scale, we attached proper weights to each token: weights 10, 5 and 2 were assigned to frequencies 1, 2 and 3 respectively. The distance of the first weight (10) and the second (5) shows the significance of the most rare words. The final score is the following:

$$\frac{\sum_{n=1}^{m} word_n(weight)}{m}, \text{ m=number of words of each tweet} \quad (1)$$

This formula looks up every word of the given tweet in our frequency dictionary. If the word is found there, it attaches a weight according to the scale we described above. For normalization purposes, we divide the weight scores with the number of words of each tweet. The descriptive statistics of this variable distribution show high concentration around the 0.4 score with a maximum value of 10. Almost half of the dataset tweets have a score of zero.

4.3 Meanings

We used the Balkanet packet of Wordnet to extract the meanings of each word, because the use of a word with multiple meanings implies ambiguity and eventually irony. For instance, the one-liner 'Change is inevitable, except from a vending machine' exploits the ambiguity, and consequently wrong expectations, induced by the word change (Mihalcea et al., 2006). Our algorithm looks up in Balkanet every word of each tweet. If a word has multiple synsets (meanings), we count their number and add them to a score. This process is repeated for every tweet. The descriptive statistics of this variable distribution show high concentration between scores 0.2 and 3 with a maximum value of 85. Almost a third of the dataset tweets have a score of zero.

4.4 Lexical

The lexical attributes of each tweet were: repeated letters, metaphor words and punctuation. The repeated letters characterize a spoken-verbal emotionally charged expression. This phenomenon is called prosody, altering the intonation of speech like singing (Cheang et al., 2008). Also, we track the occurrences of words that showcase figurative language. For the example, the word "like" in Greek is written as σαν, σα & σάν.

The punctuation feature is the aggregation of exclamation marks, question marks, dots and semicolons. The semicolon is used in Greek instead of the '?' symbol. The descriptive statistics of this variable distribution show high concentration of scores 1 and 2 with a maximum value of 5. Two thirds of the dataset tweets have a score of zero.

4.5 Emoticons

The emoticon feature detects all the possible variations of the smiley, sad and mocking faces. The existence of emoticons is a slight indication of irony, due to emotional charge. The value of the Emoticons feature is binary: 'true' if at least one emoticon appears in the tweet, 'false' otherwise.

¨Ο Καμμένος σαν λιοκαμένος μου φαίνεται'

Spoken Score	→	0
Rarity Score	→	1.67
Meaning Score	→	0
Lexical Score	→	1
Emoticon Score	→	0

Figure 1: Example tweet assigned with feature scores

The above example (Fig. 1) displays a random political Greek tweet about a party leader and translates roughly to "Kammenos looks like sunburnt". This tweet exploits a wordplay with the name of the party leader, which in greek sounds like the word "sunburnt". As a result, the lexical score is above zero, since it uses figurative speech ("like"), as well as the rarity score, expressing a colloquial term of the word sunburnt in Greek.

5. TEST AND RESULTS

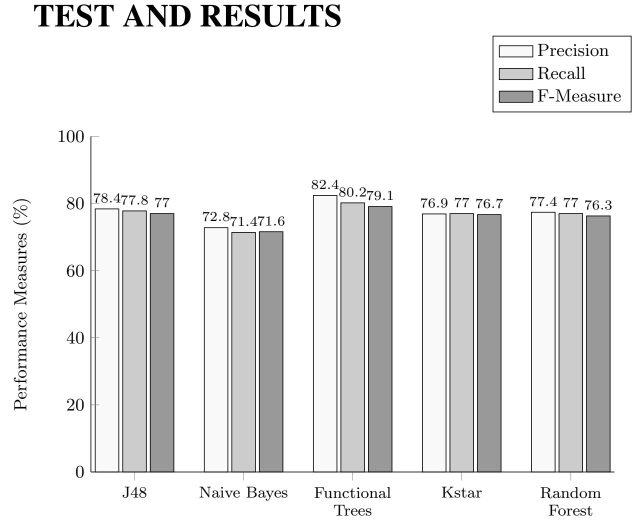

Figure 2: Performance of the learning algorithms

5.1 Training

Following the data preprocessing, we used the complete dataset including the features' scores. We labelled a small amount of tweets manually (n=126) in order to train the classifier. The distribution of the dependent variable is: 74 ironic and 52 non-ironic tweets, gathered by randomly sampling the big dataset. The resulting set was loaded on Weka

Party	Ironic	Total	Ironic/Total %	Election Results %	Fluctuation from 2009 %
ANTARSYA	511	1240	41	1.19	0.83
PASOK	1440	2240	64	13.18	-30.74
DHMAR	497	1223	41	6.11	new
KKE	504	1262	40	8.48	0.94
ND	252	560	45	18.8	-14.6
SYRIZA	2649	4371	61	16.7	12.1
XA	1040	1821	57	6.9	6.68
ANEL	280	680	41	10.6	new

Table 3: Ironic tweets that received every party before elections and their election results. The Fluctuation describes the difference between the May 2012 election results and the previous election of 2009.

and used to train on multiple algorithms according to the 10 fold cross validation technique. Apart from probabilistic algorithms, we involved decision trees as well, in order to be able to rank the significance of the features. The training algorithms with the best performance were: J48 - the Weka version of C4.5 (Quinlan, 1993), Naive Bayes (John et al., 1995), Functional Trees, KStar (Cleary et al., 1995) and Random Forests (Breiman, 2001). The best performing algorithm on average was Functional Trees (Precision=82.4). Functional Trees combine a univariate decision tree with a linear function by means of constructive induction. Decision trees created from the model are able to use decision nodes with multivariate tests, and leaf nodes that make predictions using linear functions (Gama J., 2004).

5.2 Feature Selection

The Decision Tree algorithm as described above, creates a univariate tree with linear functions on its leaves. The features **significance rank**, ordered by decreasing significance, based on Information Gain, is the following:

1. Rarity
2. Lexical
3. Emoticon
4. Meanings
5. Spoken

5.3 Prediction

The best performing classification model induced with the Functional Trees algorithm was applied to the big unlabelled dataset (n=44.500 tweets) in order to get the irony predictions of the unlabeled tweets. Due to the fact that our test dataset is unlabelled, we can't evaluate the model's validity directly. An indirect, qualitative evaluation is attempted in the following section, comparing the volume of irony in tweets with the actual election results.

5.4 Hypothesis Evaluation

After the automatic labelling of the unlabelled 'before elections' dataset, we counted the ironic percentage on every 'party' dataset. Interestingly, the actual election results are not directly correlated but there is a trend is discovered between the parties that receive irony and their election votes' percentage fluctuation retrospectively.

Table 3 consists of the sub-dataset of 'parties before elections'. One should note that the irony percent shows a trend that may be interpreted as the hype for every party. The 'Election results' column refers to the actual results of the May 2012 elections. The fluctuation(last column) describes the difference between the May 2012 election results and the previous of 2009. The fluctuation percent shows a ***trend on the edges***, so that the 'loser' parties of ND and PASOK are getting ironic tweets as well as the 'winner' parties SYRIZA and XA.

6. LIMITATIONS AND FURTHER WORK

Twitter in Greece does not have the same reach as in other countries. According to social media analysis sites, the Greek users of Twitter are 412.272 [3] (3.7% of population). The average user is usually young, well educated and liberal, something important for our political context. As a result, our findings are filtered through this demographic.

On the technical side, we did not use a stemmer or a lemmatizer, because our hypothesis is depending on rare words or wordplays which would be eliminated. Furthermore, the informal nature of the text would render the performance of such tools rather useless for a morphologically rich language such as Greek. Another restraint for our study was the shortage of tested NLP tools for the Greek language. Some available tools are not well documented or not accessible. As a result, our study is focused mainly on self-developed tools for mining the features from the text. On the semantic analysis, the meanings score was not effective due to the fact that the Balkanet framework does not support grammatical conjugation, resulting to fewer results. It accepted only the nominative case. Also, an interesting future work could be the in depth examination of the discriminating power of each of the features used/proposed. Of course, the manually trained dataset is pretty small and should be expanded for more accurate results.

7. CONCLUSION

We researched on the theoretical ground of the Twitter uses, especially on influencing and motivating the individual to criticize and joke about politics. The empirical study, attempts to detect irony on Greek political tweets, to automatically label a big unlabelled dataset of them and to seek underlying relations between the irony that the parties receive and their actual election results. The precision (82.4) on a small labeled dataset is encouraging, and feature selection confirmed our initial hypothesis of the high significance of the selected attributes on detecting irony.

The real-world application of irony detection could be useful to polling companies to get the pulse of social media in

[3]statistic from social media analysis site trending.gr

election periods as well as to the parties to get feedback. Another business-oriented aspect could be its use by brands in crisis management situations to leverage the opinion of the web.

8. REFERENCES

[1] Barbieri, Francesco, and Horacio Saggion. ”Modelling Irony in Twitter.”Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics. 2014.

[2] Bird, Steven, Edward Loper and Ewan Klein (2009), Natural Language Processing with Python. O'Reilly Media Inc.

[3] Breiman, Leo. ”Random forests.” Machine learning 45.1 (2001): 5-32.

[4] Cheang, Henry S., and Marc D. Pell. ”The sound of sarcasm.” Speech communication 50.5 (2008): 366-381.

[5] Cleary, John G., and Leonard E. Trigg. ”K*: An instance-based learner using an entropic distance measure.” Proceedings of the 12th International Conference on Machine learning. Vol. 5.1995.

[6] Colleoni, Elanor, Alessandro Rozza, and Adam Arvidsson. ”Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data.” Journal of Communication 64.2 (2014): 317-332.

[7] Davidov, Dmitry, Oren Tsur, and Ari Rappoport. ”Semi-supervised recognition of sarcastic sentences in twitter and amazon.” Proceedings of the Fourteenth Conference on Computational Natural Language Learning.Association for Computational Linguistics, 2010.

[8] Gama, João. ”Functional trees.” Machine Learning 55.3 (2004): 219-250.

[9] González-Ibánez, Roberto, Smaranda Muresan, and Nina Wacholder. ”Identifying sarcasm in Twitter: a closer look.” Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2. Association for Computational Linguistics, 2011.

[10] John, George H., and Pat Langley. ”Estimating continuous distributions in Bayesian classifiers.” Proceedings of the Eleventh conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 1995.

[11] Kalsnes, Bente, Arne H. Krumsvik, and Tanja Storsul. ”Social media as a political backchannel: Twitter use during televised election debates in Norway.”Aslib Journal of Information Management 66.3 (2014): 313-328.

[12] Kermanidis, Katia Lida, and Manolis Maragoudakis. ”Political sentiment analysis of tweets before and after the Greek elections of May 2012.”International Journal of Social Network Mining 1.3 (2013): 298-317.

[13] Liebrecht, Christine, Florian Kunneman, and Antal van den Bosch. ”The perfect solution for detecting sarcasm in tweets #not.” WASSA 2013 (2013): 29.

[14] Mihalcea, Rada, and Carlo Strapparava. ”Learning to laugh (automatically): Computational models for humor recognition.” Computational Intelligence 22.2 (2006): 126-142.

[15] Quinlan, J. Ross. ”C4.5: Programming for machine learning.” Morgan Kauffmann (1993).

[16] Rajadesingan, Ashwin, Reza Zafarani, and Huan Liu. Sarcasm Detection on Twitter: A Behavioral Modeling Approach. Diss. Arizona State University, 2014.

[17] Reyes, Antonio, and Paolo Rosso. ”Mining subjective knowledge from customer reviews: a specific case of irony detection.” Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis. Association for Computational Linguistics, 2011.

[18] Reyes, Antonio, Paolo Rosso, and Tony Veale. ”A multidimensional approach for detecting irony in twitter.” Language resources and evaluation 47.1 (2013): 239-268.

[19] Tufis, Dan, Dan Cristea, and Sofia Stamou. ”BalkaNet: Aims, methods, results and perspectives. a general overview.” Romanian Journal of Information science and technology 7.1-2 (2004): 9-43.