Irony and Sarcasm Detection in Twitter: The Role of Affective Content

Detección de Ironía y Sarcasmo en Twitter: La Función del Contenido Affectivo

Delia Irazú Hernández Farías
PRHLT Research Center, Universitat Politècnica de València, Spain
Dipartimento di Informatica, University of Turin, Italy

Department of Computational Sciences
National Institute of Astrophysics, Optics and Electronics
Luis Enrique Erro 1, 72840, Santa María Tonantzintla, Puebla, Mexico
dirazuherfa@inaoep.mx

Resumen: Tesis doctoral en Informática realizada por Delia Irazú Hernández Farías y dirigida por el Dr. Paolo Rosso (*Universitat Politècnica de València*) y la Dra. Viviana Patti (*University of Turin*) en el marco de un convenio de cotutela entre la Universitat Politècnica de València, España y la Universidad de Turin, Italia. La defensa de la tesis fue en Valencia, España el 25 de septiembre de 2017 ante un tribunal compuesto por: Horacio Saggion (*Universitat Pompeu Fabra*), Elisabetta Fersini (*Università degli Studi di Milano-Bicocca*) y Roberto Basili (*Univerità di Roma Tor Vergata*). Se obtuvo la mención internacional tras una estancia de 12 meses en la Universidad de Turin.
Palabras clave: Detección de ironía, detección de sarcasmo, procesamiento de lenguaje figurado, análisis de sentimientos

Abstract: PhD thesis in Computer Science written by Delia Irazú Hernández Farías under the supervision of Dr. Paolo Rosso (*Universitat Politècnica de València*) and Dra. Viviana Patti (*University of Turin*). This thesis was developed under a cotutelle between the Universitat Politècnica de València, Spain and the University of Turin, Italy. The thesis defense was done in Valencia, Spain on September 25, 2017. The doctoral committee was integrated by: Horacio Saggion (*Universitat Pompeu Fabra*), Elisabetta Fersini (*Università degli Studi di Milano-Bicocca*), and Roberto Basili (*Univerità di Roma Tor Vergata*). The International mention was achieved after a 12 months internship at University of Turin.
Keywords: Irony detection, sarcasm detection, figurative language processing, sentiment analysis

1 Introduction

People tend to use irony and sarcasm in social media to achieve different communication purposes. Dealing with such figurative language devices represents a big challenge for computational linguistics. The fuzzy separation between irony and sarcasm could lead to confusion. Commonly, the term "irony" is used as an umbrella term also covering sarcasm[1].

Irony is closely associated with the expression of feelings, emotions, attitudes, and evaluations toward a particular target. It allows us to convey very subjective ideas and opinions in an indirect way, going beyond the literal meaning of the words. Therefore, since irony is considered as an affective manner of communication, taking advantage of affect-related information may help to identify ironic content in social media.

This thesis aimed at investigating the role of affect-related information in irony detection. We proposed to exploit affect-based information to characterize ironic content in Twitter. A complex and multifaceted phenomenon such as irony merits to be addressed

---

[1] In what follows we shall use "irony" in the same perspective.

not only considering broad aspects of affect (such as positive and negative sentiment) but also affective information in a finer-grained perspective (taking into account different models of emotions). This research was conducted paying special attention on three main aspects:

I. We analyzed the presence of different aspects of affect in ironic utterances in order to identify potential features to characterize such phenomenon in social media. We proposed a model, called *emotIDM*, which exploits an extensive set of resources covering different facets of affect ranging from sentiment to finer-grained emotions for characterizing ironic utterances. To evaluate our model, we collected a set of Twitter corpora used by scholars in previous research, to be used as benchmarks with a two-fold purpose: to compare the performance of our model against other approaches in the state of the art, and to evaluate its robustness across different aspects related to the characteristics of the corpora. Results show that *emotIDM* has a competitive performance across the experiments carried out, validating its effectiveness.

II. Aiming to investigate the differences between tweets labeled with #irony, #sarcasm, and #not, we analyzed different facets of affective information in instances containing such labels. We find data-driven arguments suggesting that the above mentioned hashtags are used to refer different figurative language devices. The results of our analysis allow us to probe distinctions and similarities between tweets labeled as #irony and #sarcasm. Our results allow us to contribute to the assumption that there is a separation between these figurative language devices. Furthermore, bearing in mind the importance of irony detection for sentiment analysis tasks, we investigated also the differences in polarity reversal terms of such tweets.

III. Detecting irony in user-generated content could have a broad range of applications. One of them is on sentiment analysis (SA). We analyzed the impact of irony in the performance of systems dedicated to SA, observing a drop when they deal with such figurative language devices. We proposed an irony-aware sentiment analysis system that incorporates *emotIDM* in a first stage before determining the overall polarity of a given Twitter message. The obtained results are competitive with the state of the art.

2 Thesis Overview

This thesis is comprised of a compendium of research articles already published (two international journal papers, an international conference paper, and a chapter in a book) as well as unpublished content. It consist of 8 chapters that are briefly introduced below.

Chapter 2 (Hernández Farías and Rosso, 2016) presents an overview of some state-of-the-art approaches to deal with irony and sarcasm detection in social media. Furthermore, an analysis of the performance of sentiment analysis systems on the presence of ironic content is presented. In Chapter 3 (Hernández Farías, Benedí, and Rosso, 2015), we described an irony detection model exploiting surface patterns as well as some features coming from sentiment analysis. The latter being the most relevant ones, thereby giving insights into the importance of such information for detecting irony in Twitter.

Chapter 4 (Hernández Farías, Patti, and Rosso, 2016) presents *emotIDM* which considers several aspects of affect ranging from sentiment to fine-grained models of emotions. The obtained results outperform the performance of the state-of-the-art approaches validating the importance of affect-related information for irony detection in Twitter. In Chapter 5 (Sulis et al., 2016), we analyzed different facets of affective information in tweets labeled with #irony, #sarcasm and #not with the aim of distinguishing between them. We also considered the role in terms of polarity reversal of tweets containing such hashtags.

Chapter 6 (Hernández Farías et al., 2017) describes an approach for performing sentiment analysis in tweets with figurative language. We proposed a pipeline that comprises two phases: first, we exploited emotIDM for identifying irony; then, by taking advantage of several affective resources we determine a polarity degree also considering the presence of ironic content. In Chapter 7, we presented the obtained results of some further experiments and analysis carried out with the aim to enhance the research work. Finally, in Chapter 8 we draw the main conclusions of the thesis, as well as the contributions and future work.

3 Contributions

Irony is a complex mode of communication closely related to the expression of feelings.

In this thesis, we introduced the problem of detecting irony and sarcasm in social media by considering mainly the subjective intrinsic value enclosed in ironic expressions. In this line, the following contributions were made within the development of the present research:

- *Overview of irony detection and its impact on SA.* We presented a brief description of the proposed approaches in the literature together with an analysis of shared tasks regarding SA where the participating systems were evaluated on the presence of figurative language devices. Overall, there is a drop in performance of the systems. It allows validating the assumption concerning to the importance of irony detection for determining the sentiment in a text.

- *emoIDM: an irony detection model.* We proposed a novel model for identifying irony in social media by taking advantage of different facets of affective information to capture ironic intention in Twitter. Unlike other research works, we did not build our own dataset, instead we collected a set of Twitter corpora previously used in the literature. The obtained results overall outperform those from the related work validating the importance of such kind of content for irony detection.

- *Analysis on the use of different hashtags related to ironic phenomena.* We investigated the use of different hashtags (#irony, #sarcasm, and #not), concluding that these labels are indeed used to tag different phenomena. Moreover, we explored the controversial subject to separate irony from sarcasm outperforming the state of the art. With respect to #not, it seems that it is used to represent a different figurative language device, although closer to #sarcasm than #irony. Furthermore, we investigated the behaviour of tweets labeled with such hashtags in terms of polarity reversal. It seems that in tweets labeled with #sarcasm often there is a full reversal (varying from a polarity to its opposite, almost always from positive to negative polarity), whereas in the case of those tagged with #irony there is an attenuation of the polarity (mostly from negative to neutral).

- *Development of an irony-aware sentiment analysis system.* Irony is a phenomenon

having an impact on the performance of systems dedicated to calculate the overall sentiment expressed in a given piece of text. We incorporated our irony detection model in a pipeline of sentiment analysis that relies mainly on sentiment and emotional-related resources. The obtained results were competitive and serve to validate the relevance of exploiting affective related features as well as the presence of irony for determining the sentiment in a given tweet.

Acknowledgments

References

Hernández Farías, D. I., C. Bosco, V. Patti, and P. Rosso. 2017. Sentiment polarity classification of figurative language: Exploring the role of irony-aware and multifaceted affect features. In *Computational Linguistics and Intelligent Text Processing - 18th International Conference, CICLing 2017, Revised Selected Papers, Part II*, pages 46–57.

Hernández Farías, D. I., V. Patti, and P. Rosso. 2016. Irony Detection in Twitter: The Role of Affective Content. *ACM Trans. Internet Technol.*, 16(3):19:1–19:24.

Hernández Farías, D. I. and P. Rosso. 2016. Irony, Sarcasm, and Sentiment Analysis. Chapter 7. In F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, editors, *Sentiment Analysis in Social Networks*. Morgan Kaufmann, pages 113–127.

Hernández Farías, I., J.-M. Benedí, and P. Rosso. 2015. Applying Basic Features from Sentiment Analysis for Automatic Irony Detection. In *Pattern Recognition and Image Analysis*, volume 9117 of *LNCS*, pages 337–344. Springer International Publishing.

Sulis, E., D. I. Hernández Farías, P. Rosso, V. Patti, and G. Ruffo. 2016. Figurative messages and affect in Twitter: Differences between #irony, #sarcasm and #not. *Knowledge-Based Systems*, 108:132–143.