Irony detection on Twitter: a natural language processing investigation

Conference Paper · July 2013

CITATIONS
0

READS
618

3 authors:

Larissa Freitas
Universidade Federal de Pelotas
23 PUBLICATIONS **131** CITATIONS
SEE PROFILE

Aline Aver Vanin
Universidade Federal de Ciências da Saúde de Porto Alegre
14 PUBLICATIONS **66** CITATIONS
SEE PROFILE

Renata Vieira
Universidade de Évora
249 PUBLICATIONS **2,377** CITATIONS
SEE PROFILE

Some Clues on Irony Detection in Tweets

Aline A. Vanin
PUCRS, FACIN
Porto Alegre, RS, Brazil
aline.vanin@ymail.com

Larissa A. de Freitas
PUCRS, FACIN
Porto Alegre, RS, Brazil
larissa.freitas@acad.pucrs.br

Renata Vieira
PUCRS, FACIN
Porto Alegre, RS, Brazil
renata.vieira@pucrs.br

Marco N. Bochernitsan
PUCRS, FACIN
Porto Alegre, RS, Brazil
marco.bochernitsan@acad.pucrs.br

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing

General Terms

Languages, Theory

Keywords

Microposts, Twitter, irony detection

1. SUMMARY OF THE STUDY

Microposts on Twitter allow users to express ideas and opinions dynamically, although in a very limited space. The high volume of data available provides relevant clues about the judgment about certain products, events, services etc. While in sentiment analysis the common task is to classify the utterances according to their polarity, the detection of ironic statements represent a big challenge for this task. In this study, we analyze and implement some patterns that may be associated to ironic statements in Brazilian Portuguese. A common ground between the author of the tweets and their audience is required in order to establish some background information on the text; thus, contextual features, such as the specificity of a domain, the period of time, the textual support and genre (Twitter and tweets, for example), are considered.

Irony may be seen as a complex mechanism of communication that is governed by pragmatic principles, and it is often confused with sarcasm, satire or parody. In this study, we will base the task of capturing irony on a general concept for this phenomenon, since there are no consensus opinions on a rigid definition. We base the implementation of the patterns in the works of [1], [3], [2], [4]. We follow a pragmatic view, in which the context is organized as the reader is in touch with the text. This is based on information the reader knows about the textual genre (tweet format, time, domain and type of content) and on what is written. The elaboration of the patterns that would involve possible evidences of ironic messages considers the following: syntactic rules, POS tagging, some static expression, list of laughter expressions, specific punctuation, and symbolic language (like emoticons).

The expression "fim do mundo" ("end of the world") was chosen as our specific domain, and the search was made by using this exact expression in Portuguese. By the time we started collecting the corpus (from December 19th to 23rd), this was the trending topic on Twitter. This subject is a(n) (mis)interpretation of a claim in the Mayan calendar, according to which a probable date to the end of the world was foreseen by this people. Since most people could not believe in the truthfulness of this polemic fact, and this claim can be confirmed in the analysis of the corpus content, it was expected that the subject was the target of humorous and ironic comments throughout the texts[1].

Given the domain, the corpus was extracted from Twitter through a script, built in programming language Java, developed to collect tweets with the expression "fim do mundo". In order to facilitate the integration of the application Java with Twitter services, we used the library Twitter4J[2]. We limited the capture by using a geo-localization function, that is, tweets in Portuguese which were posted in the region of Porto Alegre - RS - Brazil (measure 15km from the center of the city). After that, we observed and annotated possible patterns for Brazilian Portuguese that may be related to ironic statements especifically. A list of patterns was implemented in Python, and the interface of the system was developed with library PyQT[3], and each pattern was analyzed separately in a different tab by a judge, a linguist who categorized each tweet as ironic/sarcastic (Y) or not (N).

The collection of posts by Twitter users under the domain "fim do mundo" is composed of 2.780 tweets (55.663 words)[4]. Fifteen patterns were implemented, and they were classified in seven categories, as follows: (C1) lists; (C2) exact expressions; (C3) part of speech; (C4) part of speech + exact expressions; (C5) part of speech + lists; (C6) part of speech + named entities; (C7) demostrative pronouns + named entities; and (C8) symbols (Table 1). From these patterns, 370 tweets were extracted and analyzed manually by the judge,

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

[1]Irony is the phenomenon we want to illustrate, and we intend to test and compare the patterns with other corpora in future works. We chose a specific domain to establish a context for the interpretation of the tweets.
[2]http://twitter4j.org/en/index.html
[3]http://www.riverbankcomputing.co.uk/software/pyqt/intro
[4]The collected dataset is available by contacting the authors of the study.

who observed that 269 of them were correctly identified, and this gives a 72% precision for the proposed approach[5]. In this study, no agreement measure was computed.

Table 1: List of patterns.

Categories	Patterns	Expression
C1	P1	List of Laughter Expression
	P2	List of Emoticons
C2	P3	"só que"
	P4	"sim,"
	P5	"seria"
	P6	"na boa"
	P7	"medo"|"Medo!"|"#medo"
	P8	"#ironia"|"#sarcasmo" ""#joking"|"#kidding"
C3	P9	ADV + ADV | ADJ + ADJ
C4	P10	"tão" + ADJ or "tão" + ADV
C5	P11	ADJ + List of Emoticons
C6	P12	DET + ADJ + (PREP+DET) + NE
C7	P13	Demostrative Pronouns + NE
C8	P14	...<EXPR>!
	P15	!*|?*|!*?*|?*!*

We have conducted tests using the Portuguese lexicons (SentiLex 2.0[6] adjectives) and Portuguese TreeTagger[7]. In tests, we used 16.833 SentiLex adjetives. Also, we constructed a list of laughter expressions, a list of emoticons and a list of demonstrative pronouns.

Table 2: Manual tagging results.

Patterns	Number of detected tweets	Number of tweets considered ironic	
P1	46	38	(82.60%)
P2	53	49	(92.45%)
P3	2	2	(100%)
P4	11	8	(72.72%)
P5	30	12	(40%)
P6	5	2	(40%)
P7	38	29	(76.31%)
P11	16	10	(62.50%)
P14	2	2	(100%)
P15	163	117	(71.77%)

The main results are shown in Table 2 in which we can observe that patterns related to symbolic language, such as laughter marks and emoticons, which are intrinsically linked to humorous messages, are the best hints to irony and sarcasm. Also, the use of heavy punctuation, that is, repetition of quotation and exclamation marks, leads to clues to ironic statements. Patterns related to static expressions were proven to be bad search choices, given the low output results and the generalized uses of these expressions. This conclusion does not apply to pattern P3, as "só que (não)" is culturally considered a static ironic expression. The patterns which were not found in our corpus (P8, P9, P10, P12, and P13) will be tested in future works. Although P11 does not present a significant result, we still believe − given our language users' intuition − that it would be possible to find better results in other corpora.

From the analysis of the corpus and the elaboration of the set of patterns described above, a system was developed for future studies. It allows the extraction of the patterns and manual tagging. The system has two tabs: in the first tab, called "Pattern detection", tweets are recovered according to the choice of a specific pattern; the second tab, called "Manual tagging", presents an option for selecting a tweet and manually classifying it according to the type of irony ('jocularity', 'hyperbole', 'understatement' and 'rhetorical questions').

Even though this is an initial study, we believe that we could offer relevant results in establishing those fifteen patterns for irony, since they represent a challenge in Sentiment Analysis in the sense that this phenomenon might compromise much of the results when mining opinion on certain subject. Although some of them could not be tested by using the corpus "fim do mundo", we could extract significant clues for this phenomenon of language. In future works, we will test these same patterns in other corpora, and we will also develop comparative studies. As native speakers of Brazilian Portuguese, our intuition is that they represent strong tendencies for irony/sarcasm. More tests with other domains will be run to analyze the accuracy of the established patterns. Based on the observation of the behavior of language in other corpora, it is possible that new patterns may be added to this list, as well as some of the existing ones may be excluded if they provide irrelevant results.

We intend to develop patterns with diminutive forms, for example, which are very common in Portuguese to express positive sentiments, like affect, tenderness and intimacy, but they may also have sarcastic or ironic connotations when the intention is to depreciate or insult certain entity [1]. Another topic that might enter in future discussions is the use of quotation marks, which, as suggested by Carvalho et al. [1], "are frequently used to express and emphasize an ironic content, especially if the content has a prior positive polarity". Also, we intend to implement the possibility of detecting tweets with more than one pattern.

[5]This result is similar to some existing state-of-the-art approaches, especially the one developed by Gonzalez-Ibanez et al. [2]. This positive result corroborates the efficacy of the study.

[6]http://dmir.inesc-id.pt/project/SentiLex-PT_02_in_English

[7]http://gramatica.usc.es/~gamallo/tagger.htm

2. REFERENCES

[1] P. Carvalho, L. Sarmento, M. J. Silva, and E. de Oliveira. Clues for detecting irony in user-generated contents: oh...!! it's "so easy" ;-). In *Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion*, pages 53–56, 2009.

[2] R. Gonzalez-Ibanez, 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*, pages 581–586, 2011.

[3] O. Tsur, D. Davidov, and A. Rappoport. Icwsm - a great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In *Proceedings of the Fouth International AAAI Conference on Weblogs and Social Media*, pages 162–169, 2010.

[4] A. Utsumi. Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from nonirony. *Journal of Pragmatics*, 32(12):1777–1806, 2000.