Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web

Raquel Justo [a], Thomas Corcoran [b], Stephanie M. Lukin [b], Marilyn Walker [b], M. Inés Torres [a,*]

[a] Speech Interactive Research Group, Universidad del País Vasco UPV/EHU, Depto. Electricidad y Electrónica, Fac. Ciencia y Tecnología, Sarriena s/n, 48940 Bilbao, Spain
[b] Natural Language and Dialogues Systems Lab, University of California, Santa Cruz, Department of Computer Science, 1156 N. High, SOE-3, Santa Cruz, CA 95064, USA

ARTICLE INFO

ABSTRACT

Automatic detection of emotions like sarcasm or nastiness in online written conversation is a difficult task. It requires a system that can manage some kind of knowledge to interpret that emotional language is being used. In this work, we try to provide this knowledge to the system by considering alternative sets of features obtained according to different criteria. We test a range of different feature sets using two different classifiers. Our results show that the sarcasm detection task benefits from the inclusion of linguistic and semantic information sources, while nasty language is more easily detected using only a set of surface patterns or indicators.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Dialogic language on the web in interactive forms of media such as social networks and online forums is very different than the newspaper articles or task-oriented dialogs typically studied in work on natural language processing [33,23,10,36,37,43]. Online conversation is both more informal and more subjective: users tend to express their opinions with highly subjective and often emotional language. Moreover, in many cases context is needed in order to understand what people are saying.

Only recently have large corpora of this type of subjective dialog language become available, including labeled corpora of tweets, user reviews, online conversations and chats [40,12,39,32,22]. However because the number of studies using these corpora are limited, in many cases there are few baselines establishing how difficult it is to understand user utterances in these contexts. In particular, our work draws on the recently released the Internet Argument Corpus (IAC), a publicly available corpus of online forum conversations on a range of social and political topics [40]. The IAC includes a large set of conversations from *4forums.com*, a website for political debate and discourse. This site is a fairly typical internet forum where people post a discussion topic, other people post responses, and a treelike conversation structure is created. The corpus comes with annotations of different types of social language categories including sarcastic vs. not sarcastic, nasty vs. not nasty, rational vs. emotional and respectful vs. insulting. Figs. 1 and 2 provide examples of posts and post pairs from the IAC.

We focus on two types of social dialogic language for which annotations are provided in the IAC distribution, namely SARCASM and NASTINESS. Our primary goal is simply to test how difficult it is to automatically classify sarcastic and nasty utterances using supervised learning techniques, independently of topic. Both sarcasm and nastiness are highly subjective utterance types, but previous work suggests that they are likely to differ in detection difficulty [24]. We hypothesize that nastiness is presented more overtly, using less figurative language, and requiring less world knowledge to recognize. See Figs. 1 and 2.

We present a set of supervised learning experiments on detection of sarcasm and nastiness in online dialog. We compare a range of feature sets developed using different criteria. One of our foci is to test whether it is possible to automatically obtain a set of features valuable for identifying different forms of social language in online conversations regardless of the topic, style, speaker, or affordances of the online forum. To do so, we integrate statistic, linguistic, semantic and emotional information into our features, as a richer alternative to purely statistical or syntactically motivated surface patterns. These feature sets are then used to establish a baseline for a rule-based classifier and for a Naive Bayes classifier. The unique contributions of this paper include:

- Methods for discovering an appropriate set of features for different types of social language.

* Corresponding author. Address: Depto. Electricidad y Electrónica, Fac. Ciencia y Tecnología, Universidad del País Vasco, Sarriena s/n, 48940 Leioa, Spain. Tel.: +34 94601 2715.

E-mail addresses: raquel.justo@ehu.es (R. Justo), tcorcora@soe.ucsc.edu (T. Corcoran), slukin@soe.ucsc.edu (S.M. Lukin), mwalker@soe.ucsc.edu (M. Walker), manes.torres@ehu.es, manes@ehu.es (M.I. Torres).

Category	Post or Post Pair
Sarcastic	**P1**: That's it? That was your post? To attack a source without anything backing you up? Bravo!
Sarcastic	**P2**: But...but...I just swallowed 56 zygotes so that they could vote by proxy through me for the Ripofflican(tm) party this upcoming election so they can win. Now I'm going to have all that indigestion for nothin'
Sarcastic	**P3**: OK, I get it! It's not being gay that is a sin; it's the gay sex that is a sin. You are such an expert on gay people! How about this EZ, what if two gay men live together, love each other, sleep in separate beds, never have sex but maybe give each other a little French kiss now and then? Is that OK, EZ? Here let me answer that question for you since your famous for avoiding the issue. Why yes Mr. Monster that is perfectly fine and a totally realistic thing to ask people to do! I know why you're avoiding answering the topic question directly. Believe me I am going to rip you a new one regardless (figuratively speaking.) You may as well let it all out. You know you want to. So go ahead, make my day.
Sarcastic	**P4-1**: "God" Does NOT play dice with Sexual Orientation. It is this way, because it was made this way before there was even the notion of a Bible, or Society. It is this way throughout the Animal Kingdom and remains so with us. **P4-2**: You need to experiment with the notion that you will be sexually attracted to your own sex. After you're done, come back and report your findings!
Sarcastic	**P5-1**: I simply mean a member of the species of "human." If something is a member of that species, it is a human being. **P5-2** If that is a human being, then is my kidney a human being too?
Sarcastic	**P6-1**: Let's support intrauterine civil registration. **P6-2**: Right, Cybererratus, uteruses are the property of church and state, not of the woman in whose body they are found.

Fig. 1. Sarcastic Posts and Post Pairs from `4forums.com`. Sarcastic examples were all reliably rated sarcastic: 4 or more turkers voted sarcastic, greater than 80%sarcastic yes count.

Category	Post Pair
Nasty	**P7-1**: Read what i actually wrote for a change Yank. "This is what socially inept losers do when they know their argument is ####, deny they ever made one. **P7-2**: I've asked repeatedly for an example of the evolution of a new body part in a population and all you've done is confirm the fact that you're an absolute #####."
Nasty	**P8-1**: I maintain that anyone who disagrees with this point of view is a depraved monster, a moral leper who is unworthy to be called a true human being. **P8-2**: And I maintain that your belief is dangerous to America and people like you should be deported to less civilized countries so you can truly appreciate what you had.
Nasty	**P9-1**: You can ignore the obvious question. This is what I find most evolutionists do, they ignore other branches of science that militate against their belief system, and legitimate questions raised by others are dismissed as being ignorant, or worse "religious". **P9-2**: Actually what they are really doing is ignoring silly irrelevent questions from clueless people with a religious agenda to promote.
Nasty	**P10-1**: i believe you just get angry when soemone doesn't agree with you 100%. **P10-2**: I'm not angry. In fact, I'm laughing at you right now.

Fig. 2. Nasty Post Pairs from `4forums.com`. Nastiness was annotated on a scale of −5 ... 5, with −5 being very nasty. The selected examples all had average nastiness ratings less than −2.5.

- Classification methods that explicitly consider the possibility of different forms of sarcasm such as hyperbole, understatement and irony [14,4].
- Feature sets which explicitly considers the semantic meaning and the problem of long utterances that include **both** the target category of sentences, e.g., sarcastic, **as well as** sentences not in the target category (e.g., not sarcastic).
- Comparison of ease of detection of two types of social language (sarcasm and nastiness) in an identical context.

2. Related work

Social networks, blogs, forums and many other websites allow people to share information. This social use of the web can provide valuable information to companies, which are therefore interested in opinion mining and sentiment analysis. Developing tools to select and analyze opinions is now a challenging goal for many companies. However, this information is informal and unstructured so that it is also a challenging topic for research in natural language processing. Truly understanding natural language requires computational models that can decoding the semantic meaning of utterances as well as the sentics, requiring methods that go beyond words to deal with concepts [8]. Cambria et al. [7] present a review of the evolution of research on these topics. According to their work, sentiment analysis has typically been performed over on-topic documents [7]. The review includes a discussion of the most relevant text features for sentiment classification, such as term and *n*-gram frequencies and presence, certain adjectives as indicators, as well as phrases chosen by POS patterns and sentiment lexicons. Revised methods included keyword spotting, lexical affinity, and concept-based approaches using web ontologies, as well as Bayesian inference and support vector machines as statistical classifiers. Recent techniques also include concept-level analysis for both knowledge-based [9,3] and statistical approaches [13]. However, concept-based approaches to date have mainly dealt with polarity detection [9,13]. Only irony, sometimes theorized as an utterance that assumes the opposite of the actual situation, has been addressed as playing the role of a polarity reverser [3,32]. In this situation, detection of irony is assumed to require a representation of the dialog context.