Chasing Rhetorical Cues to Sarcasm in Online Debate

Shereen Oraby and **Marilyn Walker**
Natural Language and Dialogue Systems Lab
University of California, Santa Cruz
{soraby,mawalker}@ucsc.edu

Abstract

Rhetorical questions are frequently used in social media, aimed as capturing an audience's attention or achieving a pragmatic goal, such as argumentation or persuasion. Although it has been observed that many of the uses of rhetorical questions are sarcastic, there is no large-scale computational work dedicated to distinguishing sarcastic from argumentative rhetorical questions, to our knowledge. To fill this gap, we expand on the only small existing dataset to include 2.5K rhetorical questions in debate forums. We use an LSTM model to distinguish between the classes, beating previous baselines with up to 0.75F for SARCASTIC and 0.76F for ARGUMENTATIVE, then analyze characteristic linguistic categories by class.

1 Introduction

Rhetorical questions (RQs) are prevalent in argumentative and persuasive discourse (Ilie, 1994). RQs are syntactically formulated as a question, but function as an indirect assertion (Frank, 1990), so previous work on RQs focuses on their formal semantic properties (Han, 1997).

Interestingly, while Gibbs (2000) and others mention that RQs are often used ironically (Schaffer, 2005; Ilie, 1994), there is no previous work that focuses on distinguishing sarcastic from argumentative (or other) uses of RQs at scale. Gibbs' (2000) taxonomy defining a non-unitary basis for irony is based on five categories: *sarcasm, hyperbole, understatement, jocularity,* and finally *rhetorical questions*. Likewise, Ilie (1994) points out that RQs and their responses may also be used to project irony, humor, and sarcasm.

In previous work, we use Gibbs' (2000) taxonomy of sarcasm to build a corpus of three different forms of sarcasm (general, hyperbole, and RQs) (Oraby et al., 2016) . The corpus of RQs in debate uses a heuristic to identify RQs depending on their position within a post in a dialog. The final RQ dataset includes 1,702 posts balanced between the SARCASTIC and ARGUMENTATIVE classes (851 posts per class).

We choose to focus specifically on characterizing the use of RQs in online debate. Our contributions in this work are:

- Development of a larger dedicated corpus of 2,496 RQs in debate based on a heuristic we define in previous work (Oraby et al., 2016).
- Presenting classification results using an LSTM model to classify SARCASTIC vs ARGUMENTATIVE RQs, beating the baseline.
- Analysis of the linguistic categories of RQs in debate, highlighting stylistic differences.

2 Data Collection

In order to do an in-depth study of RQs in debate, we need a reasonably large set of RQs. We consider debate posts from the Internet Argument Corpus (IAC) 2.0 (Abbott et al., 2016), and the corpus of 1,702 RQs from our previous work (Oraby et al., 2016).[1]

To gather the data, we define and apply a heuristic that gathers RQs based on the simplifying assumption that a question in a post is an RQ if the question occurs in the middle of the speaker's turn, and if it is followed by a statement, indicated by a period or exclamation point (Oraby et al., 2016).

Thus, the assumption is that when speakers follow a question with a declarative statement, they

---

[1] The only other related corpus that includes RQ annotations is the set of 583 RQs in the Switchboard-DAMSL dialog corpus (Rohde, 2006; Jurafsky et al., 1997).

Features	SARCASTIC			ARGUMENTATIVE			Average		
	P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
W2V (on RQ)	0.78	0.62	0.68	0.68	0.80	0.74	0.72	0.71	0.71
W2V+LIWC (on RQ)	0.76	0.68	0.72	0.71	0.79	0.75	0.74	0.73	0.73
W2V+LIWC (on Pre+RQ)	0.81	0.60	0.69	0.68	0.86	**0.76**	0.74	0.73	0.72
W2V+LIWC (on RQ+Post)	0.74	0.76	**0.75**	0.76	0.74	0.75	0.75	0.75	**0.75**
W2V+LIWC (on Full Text)	0.76	0.67	0.71	0.70	0.78	0.74	0.73	0.73	0.73

Table 1: Supervised Learning Results for RQs in Debate Forums

indicate that they are not sincerely seeking an answer. We use the same heuristic and annotation scheme on Amazon Mechanical Turk, showing Turkers the RQ and following statement, and using majority vote based on 3 out of 5 Turkers. on the IAC 2.0. Using this method, we expand the dataset to a total of 2,496 RQ pairs, balanced between the classes. For all experiments, annotators had above 80% agreement with majority.

3 Results and Analysis

We conduct supervised learning experiments using an LSTM model from Keras (Chollet, 2015) with TensorFlow backend (Abadi et al., 2016), dividing our data randomly into 80% training and 20% test and doing a grid-search on our training data for parameter tuning. For features, we use Google News-trained Word2Vec (Lowe et al., 2015). We experiment with adding in Linguistic Inquiry and Word Count (LIWC) scores (Pennebaker et al., 2001), and training with different levels of post-level context: pre (the RQ and what precedes it), RQ (just the RQ), and $post$ (the RQ and what follows it).[2]

Table 1 shows our experimental results for each class, as well as the averages. Results increase over our simple W2V baseline for both SARC and ARGUE when using LIWC categories, with best results coming from training using the $post$ context for SARC ($p < 0.05$ compared to the baseline), and on the pre for ARGUE. It is frequent that in the SARC posts, the speaker ends with a sharp remark or interjection like *"gasp!"* or an emoticon like *wink ;)* or *rolleyes 8-)*. The ARGUE posts are often nestled within sequences of questions or other RQ pairs, seeking to inform. Our top results of 0.75F SARC and 0.76F ARGUE improve on those from our previous work (0.70 for SARC and 0.71

1	Pre	[...] the argument I hear most often from so-called 'pro-choicers' is that you cannot legislate morality.
	RQ	**Well then what can you legislate?** *Every law in existence is legislation of morality!*
	Post	By that way of thinking, then we should have no laws. If someone kidnaps and murders your 3-year-old child, then let's hope the murderer goes free because we cannot legislate morality!

Table 2: A Sarcastic RQ in Context

for NOT-SARC) with an SVM classifier and n-gram features using 10-fold cross validation on our original smaller dataset (Oraby et al., 2016).

We investigate the data to better understand the role context plays in the classification. An example RQ in context is shown in Table 2, divided into the pre, RQ and $post$. We observe that although the RQ itself may not appear sarcastic on its own, the $post$ context makes the sarcasm much more pronounced. A qualitative analysis of the SARCASTIC vs. ARGUMENTATIVE data shows that sarcastic RQs are often followed by short statements that serve to mock, whereas the argumentative pairs are used to structure an argument or draw attention to a point.

Table 3 shows examples of the top LIWC features that surface for both classes in our experiments.[3] In Row 1, we observe that 2nd person mentions are frequent in the sarcastic debate, which aligns with previous work citing that having an "identifiable victim of irony, sarcasm, or both" is heuristic for ironic interpretation (Kreuz and Glucksberg, 1989; Katz, 1996; Gibbs, 2000). Informal words and more "verbal speech style" non-fluencies, including exclamations and social media slang (Rows 2-4), are also characteristic features.

---

[2]The length of the pre and $post$ context vary based on the length of the post, which range between 10-150 words.

[3]We find the top categories empirically by using learned feature weights (FW) on a subset of the data.

#	FW	Feature	Example
1	13.5	2nd Person	**Do you ever read headers?** *You got a mouth on you as big as grand canyon.*
2	10.9	Informal	**The hate you're spewing is palpable, yet you can't even see that can you?** *Hypocrites, ya gotta luv em.*
3	10.4	Exclamation	**Force the children to learn science?** *How obscene!!*
4	5.56	Netspeak	**To make fun of my title?** *lol, how that stings...*

ARGUMENTATIVE

#	FW	Feature	Example
5	7.29	Interrog.	**How do you know it's the truth?** *If it were definitive, I'd be more inclined [...]*
6	6.30	3rd Person Plural	**what's the difference?** *both are imposing their ideologies*
7	3.84	Differ.	**if you can't, why do you keep disputing me?** *like i said [...]*
8	3.35	Health	**When will the people press congress to take up abortion?** *It's the job of congress [...]*

Table 3: LIWC Categories in Debate

For argumentative debate, we observe that interrogatives as the most distinguishing features, where questions are phrases with *who, what, where, when, why, how*. "Differentiation" is a strong feature of argumentative forums (Row 7), relating to comparative words such as "against" and "alternatively", as well as technical jargon in Row 8 (genre-specific topics, such as abortion).

4 Conclusion

To our knowledge, this is the first study dedicated specifically to sarcasm in RQs. Our results show that sarcastic and argumentative uses of RQs are highly distinguishable using linguistic features and context. We show that SARCASTIC RQs are often informal and include 2nd person references, while ARGUMENTATIVE RQs contain technical jargon and interrogatives. Our future work includes exploring RQs in different domains, using more robust models, and developing more context features.

References

Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, Xiaoqiang Zheng. 2016. TensorFlow: A System for Large-scale Machine Learning. *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation.*

Robert Abbott, Brian Ecker, Pranav Anand, and Marilyn Walker. 2016. Internet argument corpus 2.0: An sql schema for dialogic social media and the corpora to go with it. In *Language Resources and Evaluation Conference, LREC2016.*

Shohini Bhattasali, Jeremy Cytryn, Elana Feldman, and Joonsuk Park. 2015. Automatic identification of rhetorical questions. In *Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers).*

Francois Chollet. 2015. Keras. Github, https://github.com/fchollet/keras.

Jane Frank. 1990. You call that a rhetorical question?: Forms and functions of rhetorical questions in conversation. *Journal of Pragmatics* 14(5):723–738.

Raymond Gibbs. 2000. Irony in talk among friends. *Metaphor and Symbol* 15(1):5–27.

Chung-hye Han. 1997. Deriving the interpretation of rhetorical questions. In *The Proceedings of the Sixteenth West Coast Conference on Formal Linguistics, WCCFL16.*

Cornelia Ilie. 1994. *What else can I tell you?: a pragmatic study of English rhetorical questions as discursive and argumentative acts.* Acta Universitatis Stockholmiensis: Stockholm studies in English. Almqvist & Wiksell International.

Dan Jurafsky, Liz Shriberg, and Debra Biasca. 1997. Swbd-damsl labeling project coder's manual. Technical report, University of Colorado. Available as http://stripe.colorado.edu/ jurafsky/manual.august1.html.

Katz. 1996. On interpreting statements as metaphor or irony: Contextual heuristics and cognitive consequences. *Metaphor: Pragmatics and Applications.*

Roger Kreuz and Sam Glucksberg. 1989. How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General* 118(4):374-386.

Shereen Oraby, Vrindavan Harrison, Lena Reed, Ernesto Hernandez, Ellen Riloff, and Marilyn Walker. 2016. Creating and Characterizing a Diverse Corpus of Sarcasm in Dialogue. In *Proc. of the*

SIGDIAL 2016 Conference: The 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Gren Corrado and Jeff Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality In *Advances in Neural Information Processing Systems 26*.

James Pennebaker, Martha Francis, and Roger Booth. 2001. *LIWC: Linguistic Inquiry and Word Count*.

Hannah Rohde. 2006. Rhetorical questions as redundant interrogatives. *Department of Linguistics, UCSD* .

Deborah Schaffer. —2005—. Can rhetorical questions function as retorts? : Is the pope catholic? *Journal of Pragmatics*, 37:433–600.

Reid Swanson, Stephanie Lukin, Luke Eisenberg, Thomas Chase Corcoran, and Marilyn A Walker. 2014. Getting reliable annotations for sarcasm in online dialogues. In *Language Resources and Evaluation Conference, LREC 2014*.