Sarcasm Detection

- Detecting Sarcasm is Extremely Easy ;) (Parde & Nielson 2018)
- Harnessing Context Incongruity for Sarcasm Detection (Joshi et al 2015)
- Sarcasm as Contrast between a Positive Sentiment and Negative Sentiment

Detecting Sarcasm is Extremely Easy ;) (Parde & Nielson 2018)

Gist

- Doain general sarcasm detection system
- Applied to twitter and amazon product reviews
- Contains error breakdown

Intro

- Sarcasm is difficult even for humans
  - Primariy indicated using prosodic rather than syntactic cues
- Previous approaches have been largely domain specific, this is an attempt at a general purpose sarcasm detection system

Background

- Tweets may be expecially challenging because the text limit may encourage brief coments that require more contextual information
  - The example of saying "Great" just after an election may be understandable to others at that point in time but for an automatic system that is not aware of such events, it

becomes very difficult.

- Rajadesingan et al 2015 "developed behavioral models of sarcasm usage specific to individual users" (p. 22)
- Sarcastic tweets are sampled using hashtags indicating sarcasm, Amazon reviews are sampled using star ratings
- The prior work (Parde and Nielson 2017) created a domain adaption system that was used prior to training the model, this achieved better performance "in predicting sarcasm in Amazon product reviews over models that trained on reviews alone or on a a simple combination of reviews and tweets" (p. 22)

Sarcasm detection methods

Data source

- Train
  - 3998 tweets, 1003 Amazon product reviews
- Test
  - 1000 tweets (609 non-sarcastic and 391 sarcastic)
  - 251 amazon reviews (87 sarcastic and 164 non-sarcastic)

Features

- Contains Twitter Indicator
  - "Multiple binary features indicating whether the instance contains one of th esarcasm-related has-tags, emoticons, and/or indicator phrases learned by Maynard and Greenwood (2014)" (p 23)
- "Twitter-Based predicates and situations
  - "Multiple binary features indicating whether the instance contains a positive predicate, a positive sentiment and/or negative situation phrase learned by Riloff et al. (2013)

from a corpus of tweets. Includes an additional binary feature that indicates whether one ofo those positive preedicates or sentiments precedes one of those negative situation phrases by <= 5 tokens"

- Star Rating
  - "Number of stars associated with the review" (p 23) left blank for tweets
- Laughter and interjections
  - "Multiple binary features indicatingi whether the instance contains: hahahaa, haha, hehehe, hehe,jajaja, jaja, lol, lmao, rofl, wow, ugh, and/or huh" (p 23)
- Specific characters
  - "Multiple binary features indicating whether the instance contains an ellipsis, an exclamation mark and/or a question mark" (p 23)
- Polarity
  - "Multiple features indicating the most polar (positive or negative) unigram in the instance, the polarity score (-5 to +5) associated with that unigram, the average polarity of the instance, the overall (sum) polarity for the instance, the largest difference in polarity between any two words in the instance, and the percentages of positive and negative words in the instance" (p 23)
- Subjectivity
  - "The percentages of strongly subjective positive words, strongly subjective negative words, weakly subjective positive words, and weakly subjective negative words in the instance" (p. 23)
- PMI
  - "Multiple features indicating the highest number of consecutive repeated characters in the instance (e.g., Sooooo => 5) and the higehest number of consecutive punctuation characters in the instance" (p 23)
- All-Caps
  - "Multiple features indicating the number and percentage of all-caps words in the instance" (p. 23)
- Bag of words
  - Features for words most closely associated with the different training pairs (e.g. Amazon - Sarcastic, Amazon non-sarcastic, twitter sarcastic etc.)
  - Features for most common words in each of these different class source pairings.

Classification Algorithm

Naive bayes using Daume III (2007)'s method for domain adaptation. to generate source, target and general feature mappings.

Results

.59 F-score on twitter data, 1% over previous literature (not really meaningful) Recall of system is much higher (.68 vs .62) at the cost of some precision (53 vs 55). .78 F-score on Amazon reviews, much higher than previous results (Buschmeier et al 2014) (78 to 74). Once again, much higher recall (82 to 69) at the cost of precision (75 to 85)

Error analysis

- Many did not convey sarcasm once the sarcastic hash tags were removed (23)
- 8 only had sarcastic content in the hashtags
- 13 tweets were discovered not to be sarcastic upon manual inspection
- 63 Required world knowledge to know that it was sarcastic.
- Highly negative
- Reviews also had story-like passages that were sarcastic. E.g. a narrative where the thing being reviewed is doing things that are impossible.

Harnessing Context Incongruity for Sarcasm Detection (Joshi et al 2015)

Gist

- The key part of this paper is that incongruity e.g. clashes in sentiment are central to the detection of sarcasm
- "It must be noted that our system only handles incongruity between the text and common world knowledge (i.e. the knowledge that '*being stranded*' is an undesirable situation and, hence, '*Being stranded in traffic is the best way to start my week*' is a sarcastic statement)." (p 758)
- "This leaves out an example like '*Wow! You are so punctual*' which may be sarcastic depending on situational context" (p 758)
- Explicit Incongruity is where there are polarity signifying words that make the clash in sentiment apparent
- Implicit incongruity is where there are phrases that imply a particular sentiment conventionally. **These are the ones that seem the most interesting to see how they deal with them**.

Dataset

Primarily focused on tweets.

- Tweet-A (5208 Tweets, 4170 sarcastic) Downloaded by looking for certain hash tags (#sarcasm, #sarcastic adn #notsarcastic) and then did a rough quality control check to make sure that they made sense, removing wrongly labeled examples.
- Tweet-B (2278 tweets, 506 sarcastic) manually labeled for Riloff et.al 2013. I suspect what they're doing here is trying to balance the class distributions for this since predicting sarcastic tweets using the Tweet-B dataset would be quite difficult.

Discussion board datasets

- Discussion-A (1502 discussion board posts, 752 sarcastic). Obtained from the Internet Argument Corpus (Walker et al. 2012). Manually annotated,. 752 sarc and non-sarc posts are selected randomly.

ML System

Detecting incongeruity

- Identifying phrases with implicity sentiment
- Obtained using algorithm given in Riloff et al. (2013) but extract both possible polarities for both nouns and verbs
- Keeping subsumed phrases "(i.e. `being ignored' subsumes 'being ignored by a friend')"
- Riloff et al. 2013 used these phrase as part of rules while this approach is a ML approach that uses them as features.

Features

- Unigrams
- Number of capital letters
- Number of emoticons and lol's
- Number of Punctuation marks

- Boolean feature indicating whether implicitly incongruous phrases were extracted.

Explicit Incongruity features

"""

- Number of times a word is followed by a word of opposing polarity
- Length of largest series of words with polarity unchanged
- Number of positive words
- Number of negative words
- Polarity of tweet based on words present """

Analysis

- Ran into errors with subjective things (Maybe this would be resolved if they wre able to look more closely at a user's history)
- Errors when there was incongruity but it was not within the text
- Incongruity due to numbers causes errors, here's the example they provide "*going in to work for 2 hours was totally worth the 35 minute drive*"
- Pieces of sarcastic text embedded in a larger non-sarcastic text were harder to identify.
- Politeness of sarcasm introduced difficulties.

Sarcasm as Contrast between a Positive Sentiment and Negative Sentiment

Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, Ruihong Huang

Novel bootstrapping algorithm that learns lists of positive sentiment phrases and

> "Bootstrapping algorithm that automatically learns phrases corresponding to negative sentiments and phrases corresponding to negative situations" p. 705

Bootstrapped learning of positive sentiments and negative situations

> "Our goal is to create a sarcasm classifier for tweets taht explicitly recognizes contexts that contain a positive sentiment contrasted with a negative situation" p. 706

They're learning phrases that have positive or negative connotations using a single seed word "love" and a collection of sarcastic tweets.

> "Operates on the assmption that many sarcastic tweets contain both a positive sentiment and a negative situation in close proximity, which is the source of the sarcasm" p. 706.

They focus on positive verb phrases and negative complements to that verb phrase.

They don't parse because, well, parsing tweets is messy and hard. Instead they use just part of speech tags and proximity as a proxy for syntactic structure.

> "We harvest the n-grams that follow the word 'love' as negative situation candidates. WE select the best candidates using a scoring metric and add them to a lsit of negative situation phrases. p.706

> Next we explait the structural assumption in the opposite direction. Given a sarcastic tweet that contains a negative situation phrase, we infer tha tthe negative situation phrase is preceded by a positive sentiment. We harves the n-grams that preceed the negative situation phrases as postive sentiment candidates, score and select the best candidates, and add them to the list of positive sentiment phrases" (p. 706)

Using only 175,000 tweets... Quite small for such distantly supervised stuff to work.

They use #sarcasm as indicative of the sarcastic class.

They use part of speech patterns to identify verb phrases and noun phrase.

They're scoring each candidate based upon how well they corresond with sarcasm. E.g. "we score each candidate sentiment verb phrase by estimating the probability that a tweet is sarcastic given that it contains the candidate p hrase preceeding a negative verb phrase" p. 708

and "we score each remaining candidate by estimating the probability that a tweet is sarcastic given that it contaisn the predicative expression near (within 5 words) of a negative situation phrase"

> “We found that the diversity of positive sentiment verb phrases and predicative expressions is much lower than the diversity of negative situation phrases

Makes good sense that they found this ^ However, they seem to have more stringent filtering for the positive expressions...