Intonation and Computation: Sarcasm

Julia Hirschberg
LSA 2017
julia@cs.columbia.edu

Today

- What is sarcasm?
- How is sarcasm understood in human-human communication?
- How can we use computational approaches to automatically detect sarcasm?
  - Speech
  - Text

Some Examples: Twitter

- I love going to work at 5:45am on a Saturday
- Isn’t it awesome when you keep looking for a parking for a while, give up and park far, then find 5 free parking while walking?
- Cracked, bleeding lips. I love winter.
- Always a delight when you end up in a subway car with someone singing at the top of their lungs

Some Examples: Twitter

- I love going to work at 5:45am on a Saturday
- Isn’t it awesome when you keep looking for a parking for a while, give up and park far, then find 5 free parking while walking? #sarcasm
- Cracked, bleeding lips. I love winter. #sarcasm
- Always a delight when you end up in a subway car with someone singing at the top of their lungs. #sarcasm

Defining Sarcasm

- Act of saying one thing while meaning the opposite.
- Sarcasm vs irony
- How do we understand sarcasm?
  - Psycholinguistic theories
  - Brain processes

Speech Acts Theory

- Austin (1962); Searle (1969)
- Speaking = performing various acts
  - ***Phonetic (locutionary) act:*** the act of saying something; making the sounds
  - ***Illocutionary act:*** the act performed by saying something; speaker's purpose
  - ***Perlocutionary act:*** the speech act's intending effect on another person; Addressee's subsequent action

Speech Acts, Sentence Type & Intensions

- Can use one sentence type for many illocutionary acts.
  - There are at least 5 types of sentences in English, yet we distinguish 150 common illocutionary acts!
- Can use many sentence types for one illocutionary act.
- One speech act can have several communication intentions at once.

Direct vs Indirect Speech Acts

- How do we distinguish between what speakers say and what they mean (what can be inferred by it)?

- ***Direct speech act*** **VS**
- ***Indirect speech act***
  - "a speech act that is formed indirectly by performing another" – Searle

Examples

- E.g., “Can you pass the salt.”
  - What is the direct speech act?
  - What are some possible indirect speech acts?
- E.g., “This dish really could use more salt.”
  - What is the direct speech act?
  - What are some possible indirect speech act?
  - If it was said sarcastically, what is the indirect speech act?

Gricean Conversational Implicature

- “the inferences derivable from a speaker’s utterance that a listener must construct in order to use the utterance as intended”
- Saying vs. implicating
- When you speak, there are certain implications that are assumed.

Conversational Postulates

Cooperative Principle:

"Make your contributions such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged."

- Grice's Maxims
  - Quality: Be truthful.
  - Quantity: Be informative – no more or less information than required.
  - Relation: Be relevant.
  - Manner: Avoid obscurity and ambiguity

- What happens in the case of sarcasm?
  - A maxim is flouted. Which one? Or ones?

Context & Expectations

- Brain imaging studies show that literal meaning and context are both used to interpret the true meaning of an utterance (Shamay-Tsoory, Tomer & Peretz, 2005).
  - 3 areas of the brain:
    - Left hemisphere – literal meaning
    - Frontal lobes & right hemisphere - context
    - Right ventromedial prefrontal cortex – combines literal meaning and context

Other Findings

- Depending on context and expectations, indirect speech acts do not necessarily take extra processing (e.g., Gibbs, 1986).
- Sarcasm has been shown to be more widely used on mediated communication (vs. face-to-face) (e.g., Hancock, 2004).
- Spoken indicators of sarcasm differ across cultures (e.g., Cheang & Pell, 2009).

Automatic Sarcasm Recognition: Applications

- Sentiment analysis
- Intelligent agents
- Assistive technology
- Distinguishing a speaker/writer's committed beliefs from uncommitted beliefs

How Frequent Is Sarcasm?

- Strangers vs. acquaintances
- 3% for strangers are sarcastic(Coates, 1992)
- 8% for acquaintances are sarcastic (Gibbs, 2011)
- 48 sarcastic utterances in 128 half-hour talk shows (Rockwell, 2005)
- 23% of “yeah right” (Tepperman et al., 2006)

Theories of Sarcasm Representation

- Situational disparity theory (Wilson, 2006)
  - Disparity between text and contextual info
- Negation theory of sarcasm (Giora, 1995)
  - Form of negation lacking explicit negation marker
- 6-tuple representation (Ivanko & Pexman, 2003)
  - <S,H,C,u,p,p'>
  - Speaker, Hearer, Context, utterance, literal proposition, intended proposition
- Some researchers try to leverage these to automatically recognize sarcasm

Problem Definition

- Classification task
  - Narrow scope:
    - “yeah right” (Tepperman et al., 2006)
    - Similes (Veale & Hao, 2010)
- Sense disambiguation task (Ghosh et al., 2015)
  - Distinguish literal vs. saracastic uses of words
- Sequence labeling problem (Wang et al., 2015)
  - Given a sequence of tweets in a dialogue, predict the most likely sequence for happy/sad/sarcastic

Datasets

- Short texts – Twitter
  - Manual annotation
  - Hashtag-based annotation
    - #sarcasm #sarcastic #not
    - Uses author intention
    - Large-scale corpora
  - Context-based
    - Other tweets by same author
    - Dialogue context
- Reddit posts, comments (Wallace, 2014; 2015)

Sarcasm Corpora

- Call center (Tepperman et al., 2006)
- TV show (Rakov & Rosenberg, 2013)
- Similes (Veale & Hao, 2010)
- Crowdsourced non-sarcastic versions of sarcastic data (Ghosh et al., 2015)

Approaches

- Rule based
- Machine learning

Rule-based Approach

- Sarcasm's effect on sentiment (Maynard & Greenwood, 2014)
  - Tokenize hashtags that may indicate sarcasm (e.g. #irony, #notreally
  - Reverse sentiment when sarcastic hashtag present
- Sentiment mismatch (Riloff et al., 2013)
  - Recognize positive sentiment but negative situation phrase
- Sarcastic simile detection (Veale & Hao, 2010)

- Web corpus of similes: as GROUND as VEHICLE (e.g. "as happy as a clam", "*as strong as an ox*", "*as cool as a cucumber,*" "*as dead as a doornail*")
- 18% sarcastic (e.g. "***about*** as private as a park bench", Raymong Chandler "*He looked about as inconspicuous as a tarantula on a slice of angel food "*)
- "about" similes -> 76% sarcastic
- DAL scores for sentiment labels
  - 71% of positive grounds were overall negative
  - 8% of negative grounds were overall positive

Sarcastic Simile Detection

- Ground-vehicle association: are the ground and vehicle in the same "category"?
  - E.g. "as subtle as a **freight-train**"

  -> But how can we learn word categories?
    - WordNet?
    - Word embeddings?

Machine Learning Approaches

- Context Incongruity Theory (Joshi et al., 2015)
  - Explicit – "I love being ignored"
  - Implicit – "I love this paper so much that I made a doggy bag out of it"
- Prevalence
  - 11% of sarcastic tweets have positive and negative words (~18k sarcastic tweets)

- Features used to detect incongruity
  - Lexical - ngrams
  - Pragmatic – emoticons, laughter, punctuation, capitalization
  - Explicit incongruity – number of sentiment incongruities, length of pos/neg subsequences, num pos/neg words, overall lexical polarity
  - Implicit incongruity – sentiment bearing verb and noun phrases

Context Incongruity

- Approach
  - Data: Tweet-A (5208, hashtag-labeled), Tweet-B (2278, manual labels), Internet Argument Corpus (1502 posts, manual labels)
  - Feature extraction
  - SVM
- Results (F1 measure)
  - Tweet-A: 88.7%
  - Tweet-B: 61%
  - IAC: 64%

Other Features Explored

- Pattern-based text templates beyond “about” (Tsur et al., 2010)
- Sentiment lexicon-based
- Pragmatic (emoticons, mentions)
- Ambiguity, unexpectedness, emotional scenarios
- Ngrams
- Frequency and rarity of words
- Hyperbole
- Twitter context: dialogue, profile
- Universal knowledge
- User modeling

Sarcastic or Not: Word Embeddings to Predict the Literal or Sarcastic Meaning of Words (Ghosh et al, 2015)

- Treat sarcasm detection as a word sense disambiguation task: Literal/Sarcastic Sense Disambiguation (LSSD)
  - Collect target words with literal/sarcastic sense
    - Crowdsource sarcastic tweets, asking Turkers to re-phrase to convey author's intent ("I love going to the dentist/I hate going to the dentist")
    - Align original and rephrased version to identify semantically opposite words
    - Given utterance with target word, decide whether the target word is used in its literal sense

- To collect training and test data for each target word, use tweets that contain the target + sarcasm (S), target and no sarcasm (L), and target + sentiment (L_sent)
- Use *pointwise mutual information* and word embeddings to determine likelihood of sarcasm based on the different contexts of the sarcastic and literal senses of each word

Spoken sarcasm

- Prosody
- Context
- Laughter

Spoken Sarcasm Corpora

- Acted – movies, TV shows
- Natural – SWBD, Fisher

Yeah right: Sarcasm recognition for spoken dialogue systems

Tepperman, Traum, & Narayanan (2006)

Presented by: Colette Feehan

Why look at sarcasm?

- Why is sarcasm difficult for computers?
  - Violates the Gricean principle which calls for cooperative dialogue
  - Phrases' intended meaning is contrary to its literal interpretation

Definition

- "Speech bearing a semantic interpretation exactly opposite to its literal meaning…[with] **no intent to deceive on the speaker's part**"
- Importantly: We are not trying to detect deception

Overview

- Goal
  - Look at what kinds of cues might be reliable enough to get a computer to detect sarcasm
- Used the Switchboard and Fisher corpora
- Gathered data with and without surrounding context
- Cues:
  - Prosodic
  - Spectral
  - Contextual

Cues

- **Prosodic** – F0; energy

1, 2: Average pitch in "yeah" and "right" individually, normalized by the average pitch over the whole utterance
3, 4: Duration of each word, normalized by the utterance's duration
5, 6: Average energy in each word, normalized by the average energy of the utterance
7, 8, 9, 10: The number of rising and falling frames (positive and negative pitch slopes) in each word, normalized by the total frames in that word
11, 12: The number of inter-frame changes from rising to falling pitch slope within each word, also normalized by the number of frames in that word
13, 14: The overall pitch slope from the first to last frame of each word, normalized by the whole utterance's overall pitch range
15: The overall pitch slope of the whole utterance, normalized by the whole utterance's pitch range
16, 17: The pitch range for each word, normalized by the pitch range of the whole utterance
18, 19: The energy range for each word, normalized by the energy range of the whole utterance

Cues

- **Spectral** – Mel-frequency cepstrum coefficients- Short term power spectrum of a sound
  - Used the first 12 MFCCs and their delta and acceleration coefficients
- These were used to train Hidden Markov Models and then utilized for the sarcasm detection

Cues

- Contextual

  4 kinds of ‘yeah rights’

  – Acknowledgement: (perhaps a back channel)
    - A: that’s right near Piedmont
    - B: Yeah right, right…

  – Agreement/Disagreement: not used to ground understanding
    - A: a thorn in my side: bureaucratics
    - B: yeah right, I agree

Contextual Cues

- Indirect Interpretation: anecdotal; not directed at interlocutor
  - "…We have too many pets!" I thought, "Yeah right, come tell me about it…"
- Phrase internal: When 'yeah right' is not functioning as one unit
  - A: Park Plaza, Park Suites?
  - B: Park Suites, yeah right across the street…

Objective Cues

- Laughter
- Was it a response to a question
- Position in a speaker's turn
- Surrounding Pauses
- Gender – thoughts? Men supposedly use sarcasm more

Methods

- Hand coded the switchboard corpus for these contextual cues
- Trained the computational model
- Looked at which cues improved accuracy as compared to human judges

Results

	Laughter	*Q/A*	*Start*	*End*	*Pause*	*Male*
Sarcastic	0.73	0.10	0.57	0.43	0.07	0.23
Sincere	0.22	0.20	0.94	0.48	0.19	0.51

Table 1. Probability of each contextual feature given the class of speech.

	F-measure	*accuracy*
prosodic	0.00	0.69
spectral	0.35	0.77
contextual	0.52	0.84
prosodic + spectral	0.43	0.76
prosodic + contextual	0.46	0.77
contextual + spectral	**0.70**	**0.87**
all 3	0.68	0.86

Table 2. Classification results, sorted by feature set.

Conclusions

- Only prosodic cues = worst accuracy
- All three cues = improvement
- Leaving out prosodic cues in general = best accuracy
- Laughter and pauses seemed to be good cues for sarcasm
- Initial position seemed to be a good cue for sincerity
- Gender was not a good cue

What can we do with this?

- Train computer-human dialogue systems to interact better with their human
  - Should acknowledge sarcasm as part of the repartee

“Sure, I Did The Right Thing”: A System for Sarcasm Detection in Speech

Authors : Rachel Rakov, Andrew Rosenberg
2013

Study Objectives

- Using a new acted speech corpus that is annotated for sarcastic and sincere speech
- Examine a number of features that are indicative of sarcasm : baseline of accoustic features + modeling and applying prosodic contours to the task of automatic sarcasm detection.

Tepperman et al. (2006)

- Prosodic, spectral, and contextual cues
- they concluded that prosody on its own it not enough to reliably detect sarcasm, and that a combination of contextual and spectral cues distinguishes sarcasm from sincerity most accurately.

- Formal relationship in used corpora : mixed sarcasm instances
- Annotators agreed only 76% of the time

The Daria Sarcasm Corpus

- corpus created from Daria, an animated television show that ran on MTV from 1997-2001.
- acted speech but should « lead to a more natural expression of sarcasm than traditional "acted speech.” »
- 75 sarcastic sentences and 75 sincere sentences selected (context used)

The Perceptions of Sarcastic Speech Survey

- 165 speakers listen to the 150 sentences and label them as sarcastic or sincere as a forced-choice
- Anything that achieved a “Sarcasm” label with 30% agreement or less was labeled as sincere.
- Anything that was labeled as “Sarcastic” with 72% agreement higher was labeled as sarcastic.

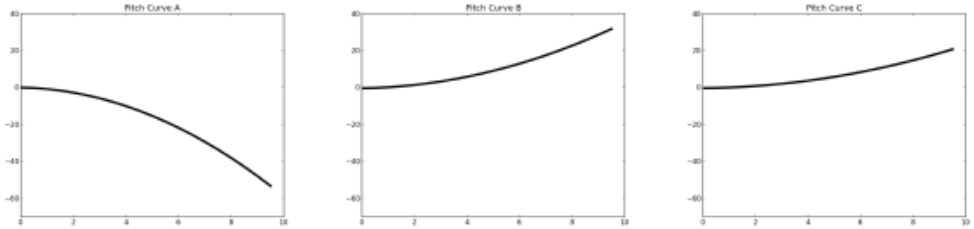
→ 112 sentences left

Sentence level features

- mean pitch
- pitch range
-standard deviation of pitch
-mean intensity
- intensity range
- speaking rate

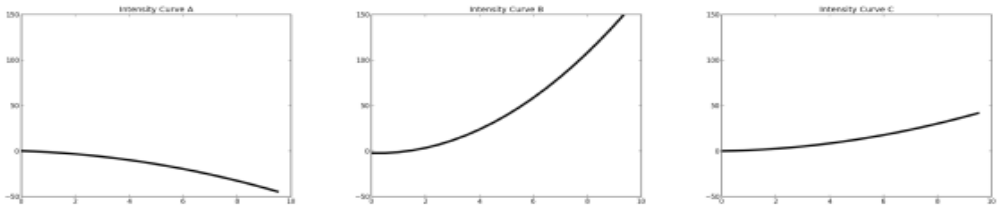
Word level features

- Pitch and intensity contours were modeled + k-means clustering algorithm



A:[-0.18,-0.26,-0.38] B:[-0.27,0.24,0.22] C:[-0.23,0.13, 0.15]

Figure 2: *Pitch contours*



A:[0.01,-0.88,-0.27] B:[-1.55,-0.81,1.22] C:[0.17,0.47, 0.28]

Figure 3: *Intensity Contours*

Word level features

unigram sequences : percentage of each curve across the sentence

- pitch unigrams A, B, and C – percent of the sentence that is modeled by word level pitch contours A, B, and C,
- intensity unigrams A, B, and C – percent of the sentence that is modeled by word level intensity contours A, B, and C,
- pitch bigram perplexity under both the sarcasm and sincere models
- intensity bigram perplexity under both models

Results

5.1.1. Feature set	5.1.2. Percent accuracy
Majority baseline	55.26
Baseline (acoustic features)	65.78
+ pitch unigrams	76.31
+ intensity unigrams	76.31
+ all unigrams	78.94
+unigrams+ pitch bigrams	76.31
+unigram+ intenbigrams	**81.57**
+unigrams + all bigrams	76.31

Table 1: *Results of experiments w/ SimpleLogistic*

Predictive Features	InfoGain
pitch range	0.20
pitch unigram a	0.17
pitch unigram c	0.12
intensity unigram c	0.09

Table 2: *Predictive Features*

- when sarcasm :

- reduced pitch range
- fewer instances of falling pitch (A) and shallow pitch rise (C)
- more instances of shallowly rising intensity (contour C)

→ sarcasm is a subtle process so abrupt shifts in intensity seem intuitively unlikely

Sarcasm & Computer-Mediated Communication

- What would we expect to see and why?
  - In light of the cooperative principle and the idea of common ground…
  - Hancock (2004)

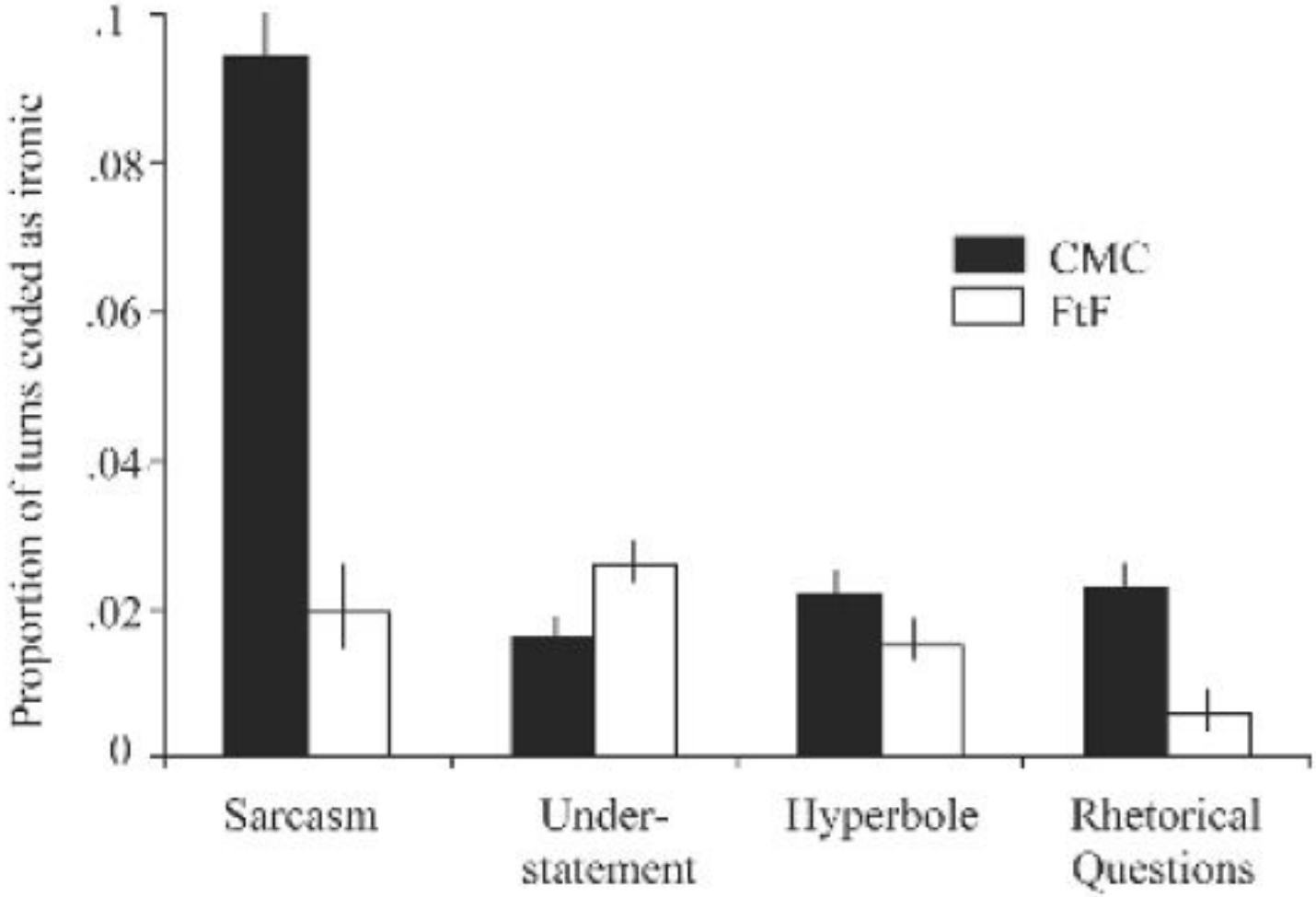


Figure 1. **Mean (Standard Error) Proportion of Ironic Turns Produced in FtF and CMC Settings.**

Additional Findings:

- Speakers in CMC used fewer cues to mark irony than in F2F.
  - F2F: prosody, laughter, hyperbolic amplifying cues
  - CMC: punctuation (not emoticons)
- Participants expressed comprehension of irony less often in CMC than in F2F.
  - Yet rated their partners as being more humorous!

- Next class: Mental Illness diagnosis