Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon (SASI)

Dmitry Davidov

Oren Tsur

Ari Rappoport

“Oooh, a sarcasm detector. That's a REAL useful invention!”

“Oooh, a sarcasm detector. That's a useful invention!”

- Model the use of sarcasm – how/why (cognitive).
- Review Summarization.
- Brand monitoring.
- Personalization of content recommendation (reviews,blogs etc.)
- May help with light autism and Asperger syndrome.

Motivation (1)

- Improve review summarization systems.
  - Identify features (size/weight, zoom, battery life, pic quality…)
  - Identify sentiment and polarity of sentiment for each feature (great battery life, insufficient zoom, distortion close to boundaries, small (keyboard/phone), “read the book”.
  - Average the sentiment for each feature

“Perfect size, fits great in your pocket” + “got to love this pocket size camera, you just need a porter to carry it for you” = ?!

Common Definition

“The activity of saying or writing the opposite of what you mean in a way intended to make someone else feel stupid or show them that you are angry. “

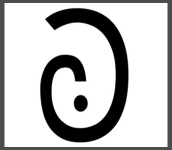
Examples

- *“Great for insomniacs.” (book)*
- *“Just read the book.” (book/movie review)*
- *“thank you Janet Jackson for yet another year of Super Bowl classic rock!”*
- *“Great idea, now try again with a real product development team.” (e-reader)*
- *“make sure to keep the purchase receipt” (smart phone)*

The challenge

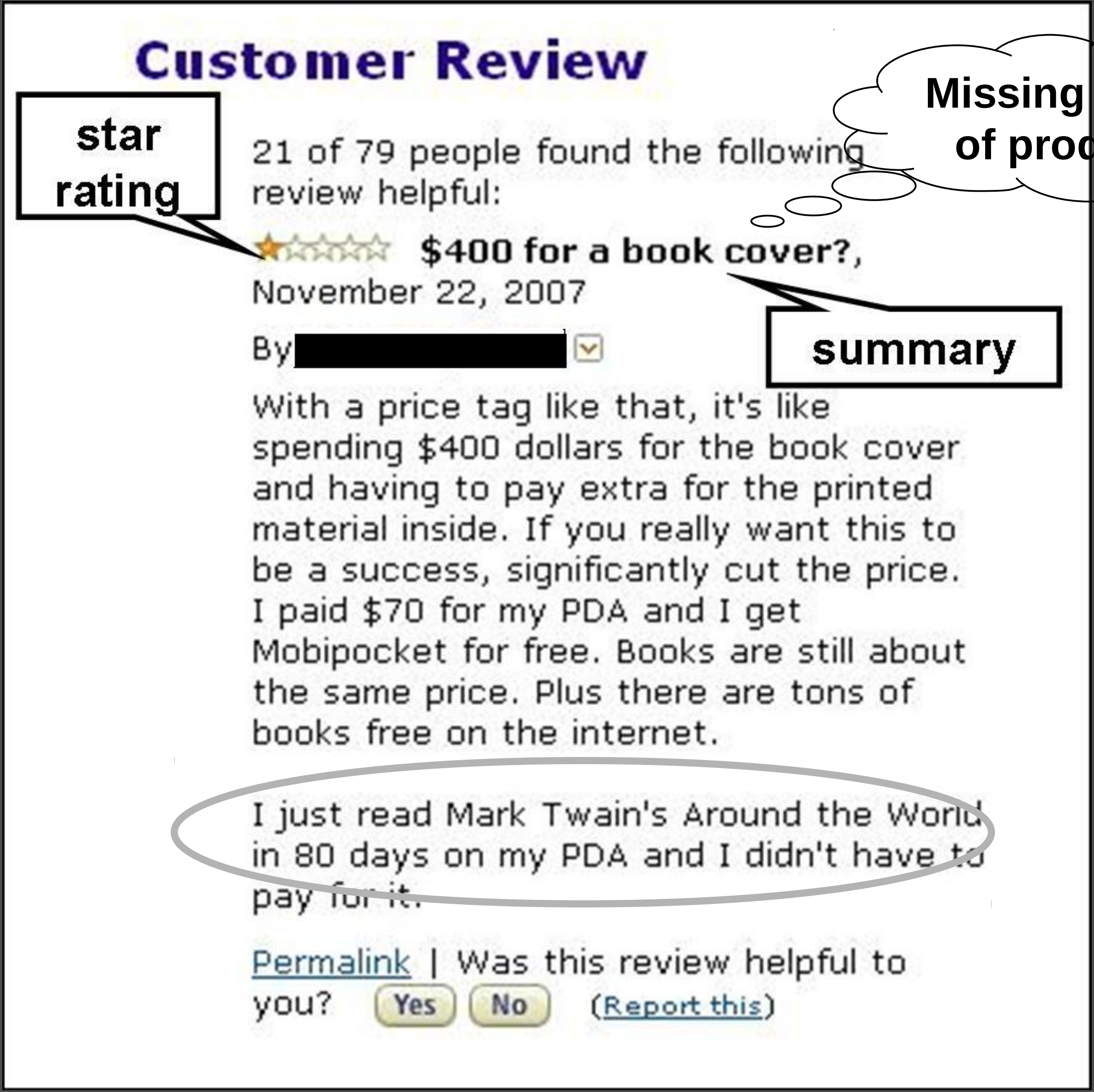
- Context.
- World knowledge.
- Missing cues in written texts.
- Hard to define.
- Violation of Grice’s maxims (?)
- Even humans sometimes fail to get sarcasm.

How do people cope?

- Temherte slaq (Some Ethiopic languages): ¡
- Mirrored question mark: ⸮
- Karl Marx in Des Kapital: [!]
- Twitter hasgtag: #sarcasm
- SarcMark 

Data

- Amazon product reviews (~66,000)
  - Books (fiction, non fiction, children)
  - Electronics (mp3 players, digital cameras, mobiles phones, GPS devices,…)
- ~6 million tweets.



Customer Review

21 of 79 people found the following review helpful:

★☆☆☆☆ **$400 for a book cover?**, November 22, 2007

By [redacted]

With a price tag like that, it's like spending $400 dollars for the book cover and having to pay extra for the printed material inside. If you really want this to be a success, significantly cut the price. I paid $70 for my PDA and I get Mobipocket for free. Books are still about the same price. Plus there are tons of books free on the internet.

I just read Mark Twain's Around the World in 80 days on my PDA and I didn't have to pay for it.

Permalink | Was this review helpful to you? Yes No (Report this)

star rating

summary

Missing name of product.

- 140 characters
- Free text
- Lacking context
- Hashtags: #hashtag
- URL addreses
- References to other users: @user

“great program at *#ISCOL*: http://www.cs.tau.ac.il/~nachum/iscol10/, @OrenTsur jabbers about twitter.”

Star Sentiment Baseline (Amazon)

- “Saying or writing the opposite of what you mean...”
  - Identify unhappy reviewers (1-2 stars)
  - Identify extremely-positive sentiment words (Best, exciting, top, great, …)
  - Classify these sentences as sarcastic.

SASI:
Semi-supervised Algorithm for Sarcasm Identification

- Small seed of sarcastic-tagged sentences. Tags 1,...,5:
  - 1: not sarcastic at all
  - 5: clearly sarcastic

SASI: outline

- Extract features from all training sentences.
- Represent training sentences in a feature vector space.
- Features:
  - Pattern based features
  - Punctuation based features
- Given a new sentence: use weighted-kNN to classify it.
  - Majority vote (over k>0)

Hashtag classifier

- #sarcasm hashtag
- Not very common

- Use this tag as a label for supervised learning.

Preprocessing

- [author],[title], [product], [company]
- [url], [usr], [hashtag]
  - *“Silly me , the Kindle and the Sony eBook can’t read these protected formats. Great!”*
  - *“Silly me , the Kindle and the [company] [product] can’t read these protected formats. Great!”*

Pattern based features

- Davidov & Rappoport 2006, 2008
- High Frequency Words (>0.0001)
- Content Words (<0.001)

- Pattern: ordered sequence of high frequency words and slots for content words.
- Restrictions:
  - 2-6 HFW
  - 1-5 slots for CW
  - Minimal pattern: [HFW] [CW slot] [HFW]

Pattern extraction from the training (seed)

“Garmin apparently does not care much about product quality or customer support”

- [company] CW does not CW much
- does not CW much about CW CW or
- not CW much
- about CW CW or CW CW.

Weights of pattern based features

- 1 : exact match.
- α : sparse match – extra elements are found between components.
- $\gamma \cdot \frac{n}{N}$: incomplete match – only *n* of *N* patterns components are found.
- 0 : no match.

“Garmin apparently does not care much about product quality or customer support”

- [company] CW does not CW much : exact match: 1
- [company] CW not: sparse match: 0.1
  - – Insertion of the word does
- [company] CW CW does not: incomplete match: 0.08
  - – One of five components (the CW) is missing: 0.1*4/5=0.08

Punctuation based

Number of !

Number of ?

Number of quotes

Number of CAPITALIZED words/letters

Classification: weighted-kNN

- For each candidate vector *v* in the test set:
- Find the k (=5) closest vectors in the training set.

The label of v is the *normalized weighted average* of $v_{1...5}$

$$Count(l) = Fraction\ of\ vectors \in training\ set\ with\ label\ \ l$$

$$Label(v) = \frac{1}{k} \sum_i \frac{Count(Label(t_i)) \dot{}\, Label(t_i)}{\sum_j Count(label(t_j))}$$

Experiments (1|2)

- 5-fold cross validation on the training set.
- Testing contribution of different features:
  - Patterns
  - Punctuation
  - Self training
  - combinations

Cross validation results: Amazon - seed; Twitter - *#sarcasm*

Amazon	Precision	Recall	F-Score
punct	0.256	0.313	0.281
patterns	0.743	0.788	0.765
punct+pat	0.868	0.763	0.812
SASI	0.912	0.756	0.827

Twitter	Precision	Recall	F-Score
punct	0.259	0.26	0.259
patterns	0.765	0.326	0.548
punct+pat	0.798	0.356	0.505
SASI	0.727	0.436	0.545

- “@USER it was sarcasm”

- “Can't wait to get home tonite”
- “Can't wait to get home tonite #sarcasm”

Gold Standard evaluation (2|2)

- Human annotation of classification of new sentences.
  - 90 sentences identified as sarcastic.
  - 90 sentences identified as non sarcastic.
  - Each sentence tagged by 3 human annotators.

Human evaluation

	Precision	Recall	F-score
star-sentiment	0.5	0.16	0.242
SASI (AM)	0.766	0.813	0.788
SASI (TW)	0.794	0.863	0.827

Some nice results

- *“If you are under the age of 13 or have nostalgia for the days when a good mystery required minimal brain effort then this Code’s for you”*
- *“@USER and Nicole rode their bikes to church... Now it looks like It's going to rain. Great!”*
- *“Dear iPod: why can't you read my mind and play the music i want to hear???”*
- *“thanks but no thanks i will NOT be checking them out today or EVER for that matter. not unless i want my ears to bleed.*

Thank you! really.[*]

[*] honestly