

Opinion mining and sentiment analysis

Bo Pang and Lillian Lee, 2008

Presented by: Jeffrey Bigg

Goals for presentation

- To distill the contents of this 135 page survey paper to material relevant for this course
- To show off the main problems associated with opinion mining (No specific algorithms!)
- To suggest how topics from this course can be applied to improve the quality of opinion and sentiment analysis

Motivation

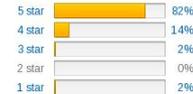
- People do still care about what other people think
- Companies are concerned with product reviews from customers, but would rather not spend extra resources analyzing unstructured text data
- Opinions can help determine the pricing of products
- Users also are concerned with product reviews, and want to do research on products without being overwhelmed

Examples: The Making of 40 Photographs > [Customer Reviews](#)

Customer Reviews

★★★★☆ 49

4.7 out of 5 stars ▾



[Examples: The Making of 40 Photographs](#)

by Ansel Adams

Format: Paperback | [Change](#)

Price: **\$30.50** + Free shipping with Amazon Prime

Rate this item



[Write a review](#)

Top positive review

[See all 47 positive reviews >](#)

168 people found this helpful

★★★★★

How Did You Make That Photograph, Mr. Adams?

By Donald Mitchell on December 21, 2000

An essential book for all photography fans!

In 1983, Ansel Adams picked 40 of his most memorable and diverse black and white photographs as examples of his work. For each one he wrote a brief essay that described the circumstances of deciding to photograph the subject, how he came to prepare for the photography, his companions, special challenges that occurred along the way, how he selected the composition, tricky light and shadow conditions encountered, technical details of how the image was captured (equipment, film speeds, settings, filters, lenses, etc.), technical details of printing the image, and the surprises he experienced.

In the midst of all this, he shares his philosophy of life, nature, and the art of photography. It's like attending a master class with a genius. Even if you know nothing about photography, this book will open your eyes to new ways of

[Read more](#)

Top critical

[See all 2 critical](#)

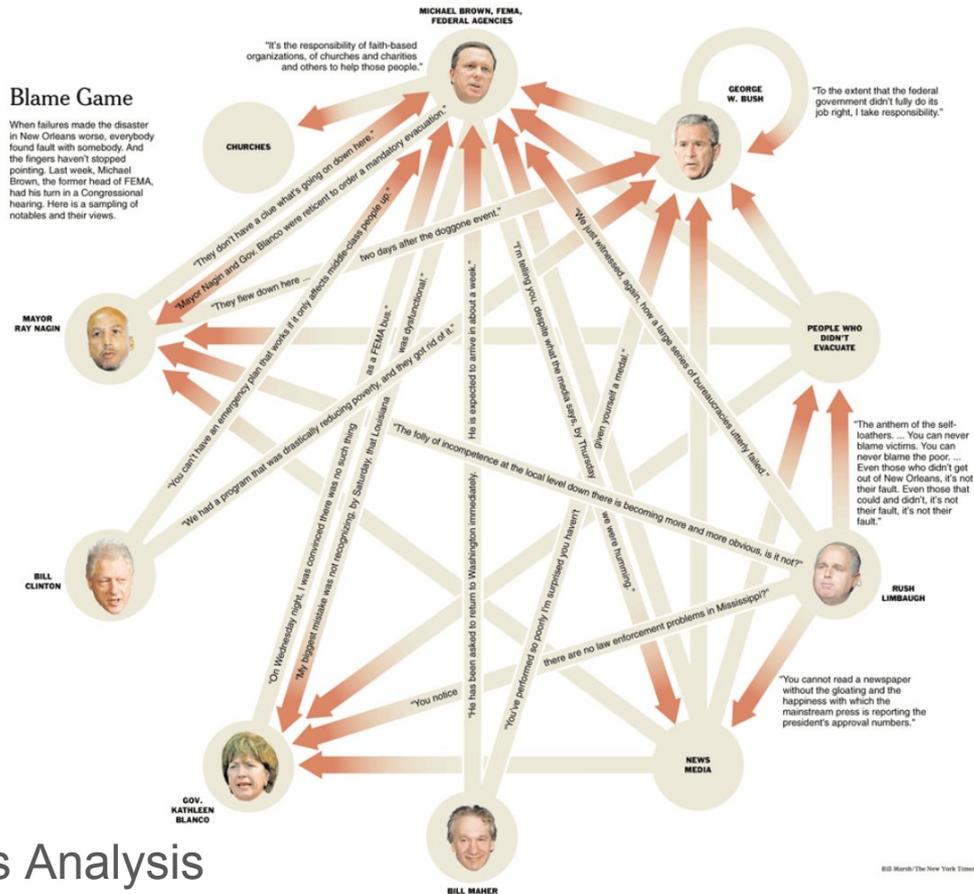
★☆☆☆☆

By Sentinel7 on

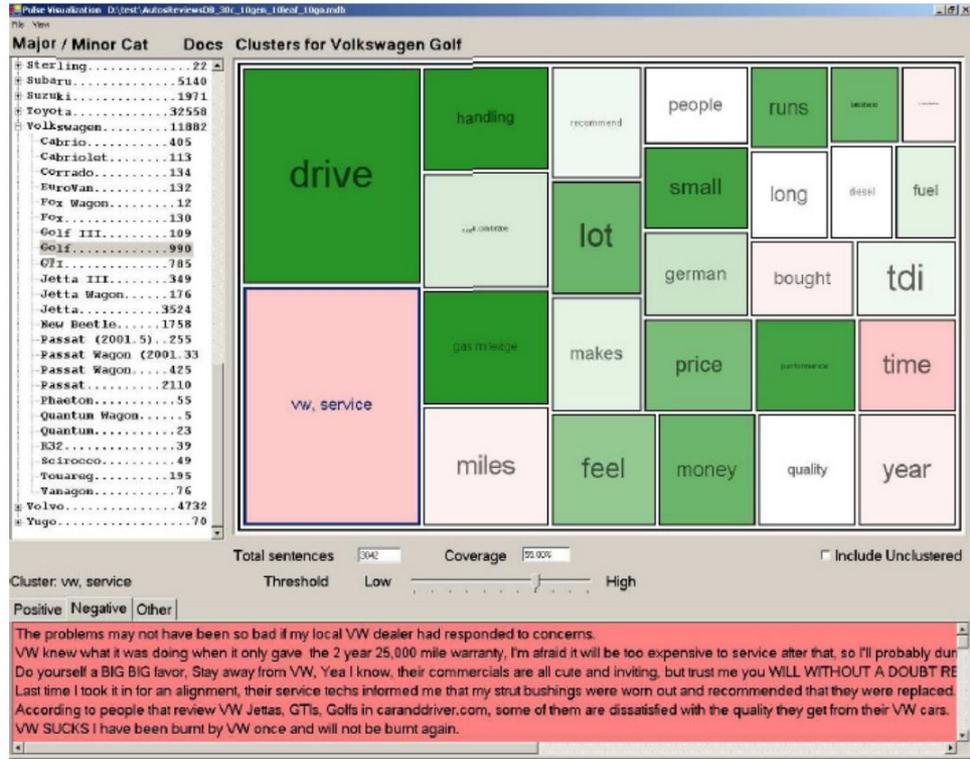
I am sure it is a

Blame Game

When failures made the disaster in New Orleans worse, everybody found fault with somebody. And the fingers haven't stopped pointing. Last week, Michael Brown, the former head of FEMA, had his turn in a Congressional hearing. Here is a sampling of notables and their views.



The New York Times Analysis



Multi Document review summarization

Things that we have seen in this course

- Phrase mining + Topic mining
 - KERT
 - ToPMine
 - SegPhrase
- Entity typing
 - ClusType

Opinion mining can generally be seen as a more basic formation of entity typing of linking phrases with two main types [positive,negative] as well as degrees of sentimentality like soft assignments to each type cluster

So we're done now, right?

WRONG

Things that make opinion mining difficult

Similarly to entity mining

1. Relationship words/phrases can be hard to classify correctly
2. How to determine degrees of positivity/negativity
3. Can be very domain dependant
4. Sentiment can be context sensitive

What we say and what we mean can be two very different things.

“Sarcasm is *great*.”

“This film should be *brilliant*. It sounds like a *great* plot, the actors are *first grade*, and the supporting cast is *good* as well, and Stallone is attempting to deliver a *good* performance. However, it can't hold up.”



SPEED
LIMIT

35

UNLESS, OF COURSE,
MR. IMPORTANT IS
RUNNING LATE

Challenges to be conquered

- Rating inference
 - How positive or how negative is a review
 - Typically done using machine learning
- Agreement detection
 - Determining whether a pair of texts have the same or different sentiment labels
 - Useful in determining rating inference
- Subjectivity detection
 - When are opinions subjective or objective
 - Difference between a mediocre opinion between positive/negative and no opinion
- Joint topic-sentiment analysis
 - Using topic mining to separate opinions associated with each

There is a good chance of high error rates using these!

Rating Inference: Possible features to use

1. Term presence vs. frequency
 - a. Low frequency relation terms might suggest emphasis (example: “bugfested”)
2. Phrases can be very useful (Considered before newer phrase mining)
3. Parts of speech
 - a. NLP to extract relationships between phrases
4. Negation terms
 - a. “I like this book” vs. “I don’t like this book”

Important in many cases to have sentence level granularity due to important parts of context. Many times have the links in a sentence labeled based on the classification of the sentence.

Rating Inference: Approaches for classifying relationships

- NLP from labeled data to detect relationships
 - Determines relationship phrases and can help categorize links
- Using topic-sentiment interaction to segment relationships
 - Perform topic mining first before doing opinion mining
- Unsupervised lexicon induction to classify relationship phrases
 - Information propagation given constraints
 - relationship phrases are classified with conjunctions. (example: “elegant but over-priced”)
 - Requires some user supplied labels
- Agreement detection
 - Using this as links between words just as with lexicon induction

Objective vs. Subjective Classification

- Need to discover objective vs. subjective relationships
 - Heuristic: homophily in objective vs. subjective relationship based on locality
 - Sentences are labeled as objective or subjective, and the relationship described within are treated as such
- Relationships between discourse participants
 - Political blogs can often quote blogs of a variety of political views, and not necessarily support or agree with the blogs that are quoted
- In general: Very similar to entity typing
 - Sentiment can often change based on your point of view (ex: Republicans might view their candidate for president positively, but Democrats might view them negatively)
 - Tends to require heuristics between different types to help determine the type and degree of the opinion

IT'S A
PROVEN FACT
NOW!

I DON'T
BELIEVE IN
YOUR FACTS!



FACTS ARE FACTS!
NOT A
BELIEF.

THAT'S YOUR
OPINION!



IT'S NOT AN
OPINION.
IT'S A FACT.

FROM YOUR
SIDE!



THERE ARE
NO SIDES!
JUST FACTS!

BUT I
DON'T
THINK SO!



JUST BECAUSE YOU DON'T
THINK IT'S A FACT DOESN'T
MEAN IT ISN'T!

WELL, I HAVE
MY SUSPICIONS!



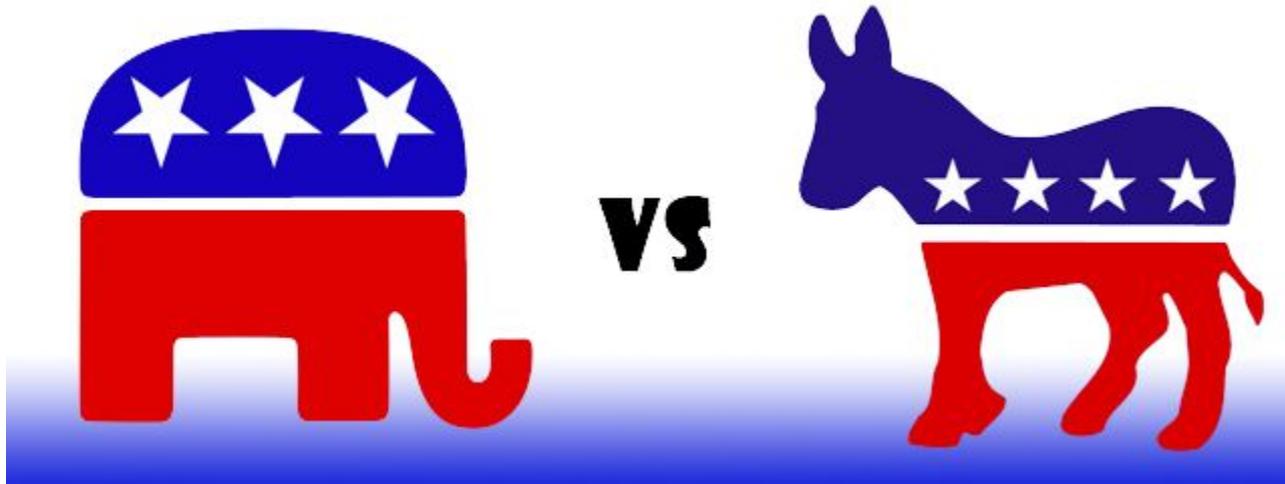
ARGH!! IT'S USELESS
ARGUING WITH YOU!!

NOW, THAT'S
A FACT!



Joint Topic Sentimentality Analysis

- Appears in 3 paragraphs in the whole paper
- General heuristic: Relationships between topics can help determine the sentimentality of relationships between different entities
- Example:



Conclusions

- Largely done as a machine learning and information propagation task
- Could be improved with quality phrase and topic modeling
- Three potential strategies could be used for future work:
 1. Strategy 1: A Priori assignment of sentimentality to linking phrases
 2. Strategy 2: Simultaneous learning of entity links with sentimentality associated with them as an EM algorithm
 3. Strategy 3: Post entity typing and link analysis assignment of sentimentality