

Sentiment Analysis

What is Sentiment Analysis?

Positive or negative movie review?



- unbelievably disappointing



- Full of zany characters and richly applied satire, and some great plot twists



- this is the greatest screwball comedy ever filmed



- It was pathetic. The worst part about it was the boxing scenes.

Google Product Search



HP Officejet 6500A Plus e-All-in-One Color Ink-jet - Fax / copier / printer / scanner

\$89 online, \$100 nearby ★★★★★ 377 reviews

September 2010 - Printer - HP - Inkjet - Office - Copier - Color - Scanner - Fax - 250 sh

Reviews

Summary - Based on 377 reviews



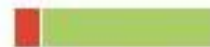
What people are saying

ease of use



"This was very easy to setup to four computers."

value



"Appreciate good quality at a fair price."

setup



"Overall pretty easy setup."

customer service



"I DO like honest tech support people."

size



"Pretty Paper weight."

mode



"Photos were fair on the high quality mode."

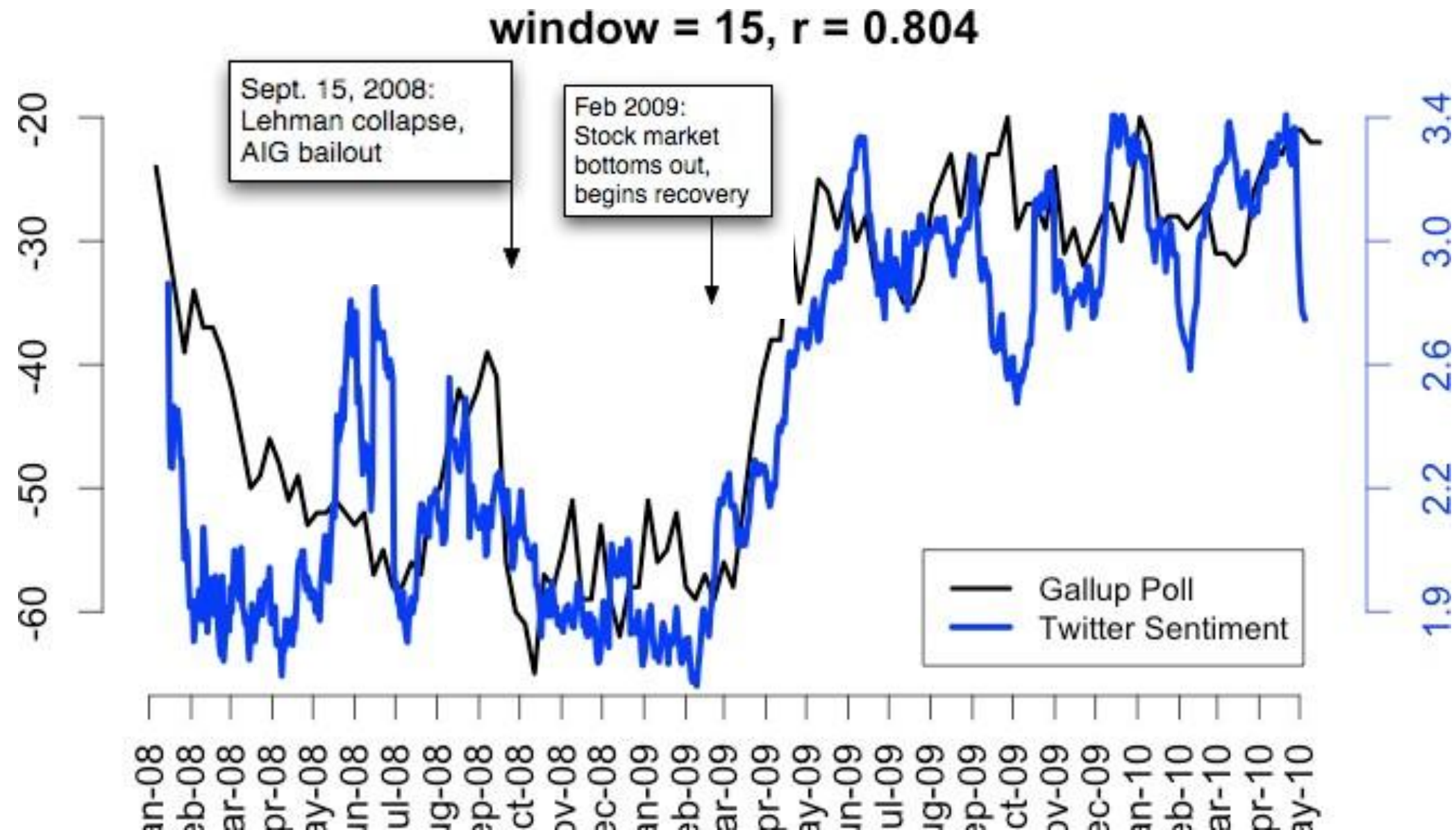
colors



"Full color prints came out with great quality."

Twitter sentiment versus Gallup Poll of Consumer Confidence

Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010.
From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In ICWSM--2010

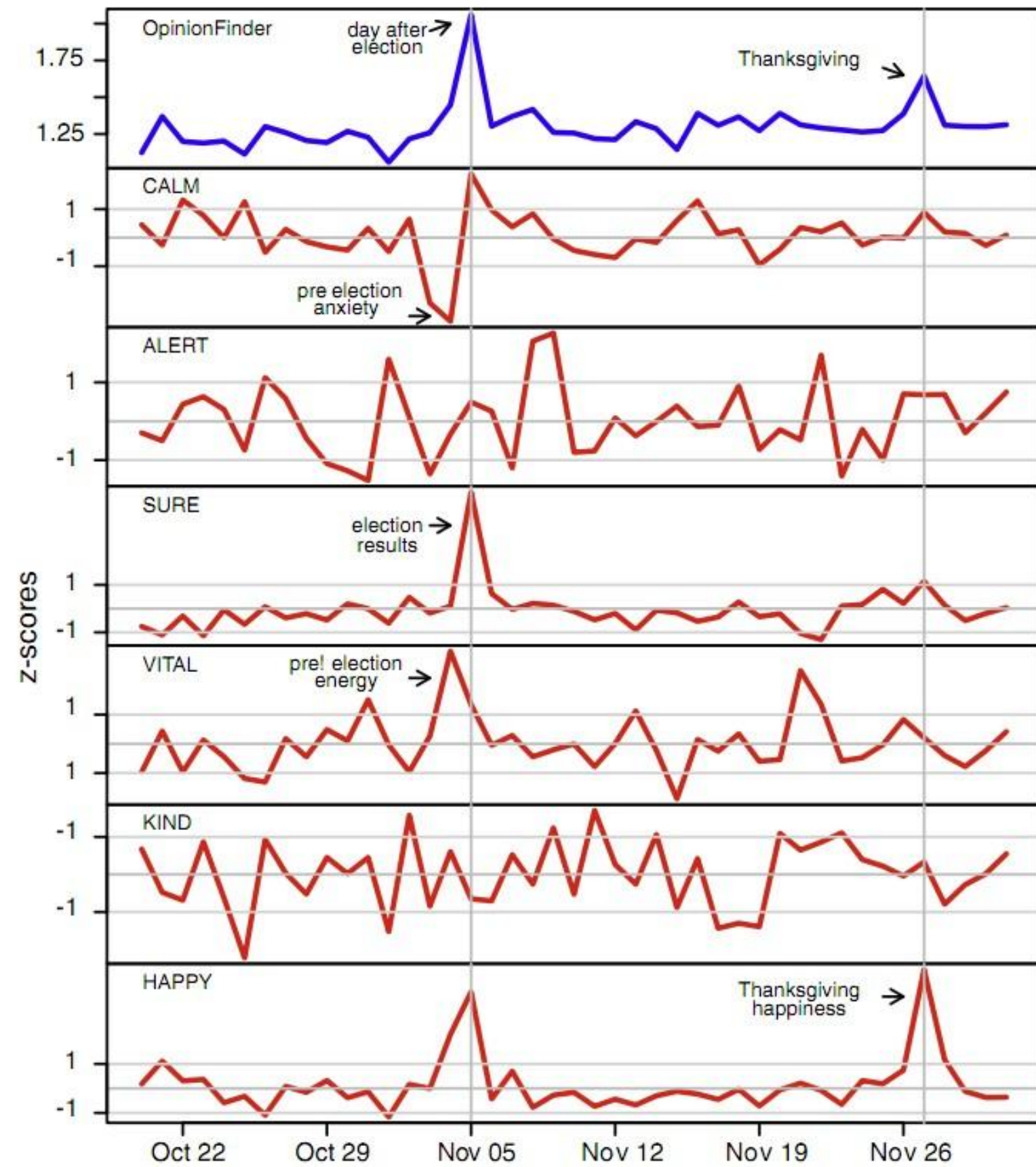


Twitter sentiment:

Johan Bollen, Huina Mao, Xiaojun Zeng.
2011.

Twitter mood predicts the stock market,

Journal of Computational Science 2:1,
1--8. 10.1016/j.jocs.2010.12.007.



Target Sentiment on Twitter

- Twitter Sentiment App
- Alec Go, Richa Bhayani, Lei Huang. 2009. Twitter Sentiment Classification using Distant Supervision

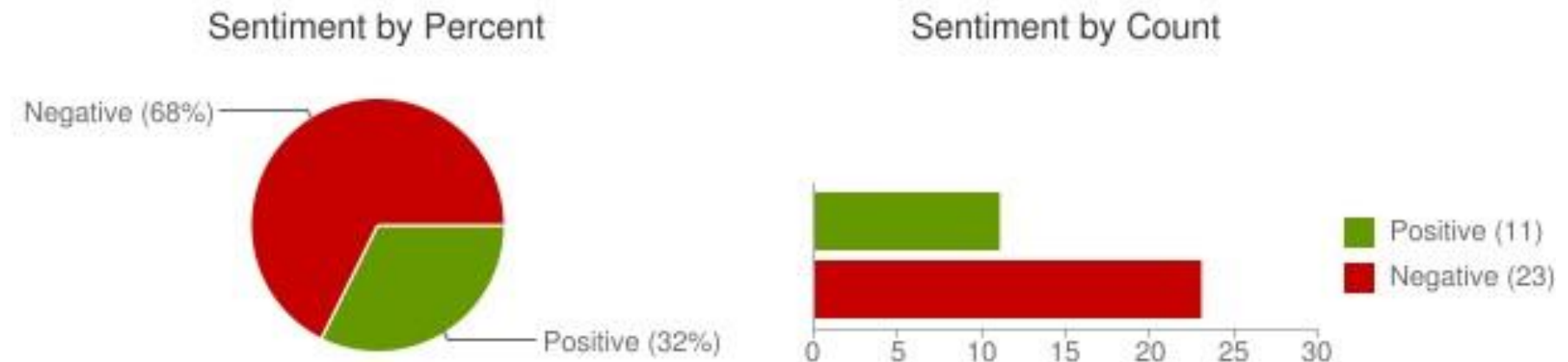
Type in a word and we'll highlight the good and the bad

"united airlines"

Search

[Save this search](#)

Sentiment analysis for "united airlines"



jljacobson: OMG... Could @United airlines have worse customer service? W8g now 15 minut
Posted 2 hours ago

12345clumsy6789: I hate United Airlines Ceiling!!! Fukn impossible to get my conduit in this d
Posted 2 hours ago

EMLandPRGbelgiu: EML/PRG fly with Q8 united airlines and 24seven to an exotic destination
Posted 2 hours ago

CountAdam: FANTASTIC customer service from United Airlines at XNA today. Is tweet more
Posted 4 hours ago

Vaccine Sentiments from Twitter



Pro-vaxxers

support and promote the benefits of vaccines



Anti-vaxxers

believe that vaccines do more harm than good

Sentiment analysis has many other names

- Opinion extraction
- Opinion mining
- Sentiment mining
- Subjectivity analysis

Why sentiment analysis?

- *Movie*: is this review positive or negative?
- *Products*: what do people think about the new iPhone?
- *Public sentiment*: how is consumer confidence? Is despair increasing?
- *Politics*: what do people think about this candidate or issue?
- *Prediction*: predict election outcomes or market trends from sentiment

Scherer Typology of Affective States

- **Emotion:** brief organically synchronized ... evaluation of a major event
 - *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood:** diffuse non--caused low--intensity long--duration change in subjective feeling
 - *cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Interpersonal stances:** affective stance toward another person in a specific interaction
 - *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*
- **Attitudes:** enduring, affectively colored beliefs, dispositions towards objects or persons
 - *liking, loving, hating, valuing, desiring*
- **Personality traits:** stable personality dispositions and typical behavior tendencies
 - *nervous, anxious, reckless, morose, hostile, jealous*

Scherer Typology of Affective States

- **Emotion:** brief organically synchronized ... evaluation of a major event
 - *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood:** diffuse non--caused low--intensity long--duration change in subjective feeling
 - *cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Interpersonal stances:** affective stance toward another person in a specific interaction
 - *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*
- **Attitudes:** enduring, affectively colored beliefs, dispositions towards objects or persons
 - *liking, loving, hating, valuing, desiring*
- **Personality traits:** stable personality dispositions and typical behavior tendencies
 - *nervous, anxious, reckless, morose, hostile, jealous*

Sentiment Analysis

- Sentiment analysis is the detection of **attitudes**
“enduring, affectively colored beliefs, dispositions towards objects or persons”
 1. **Holder (source)** of attitude
 2. **Target (aspect)** of attitude
 3. **Type** of attitude
 - From a set of types
 - *Like, love, hate, value, desire, etc.*
 - Or (more commonly) simple weighted **polarity**:
 - *positive, negative, neutral*, together with *strength*
 4. **Text** containing the attitude
 - Sentence or entire document

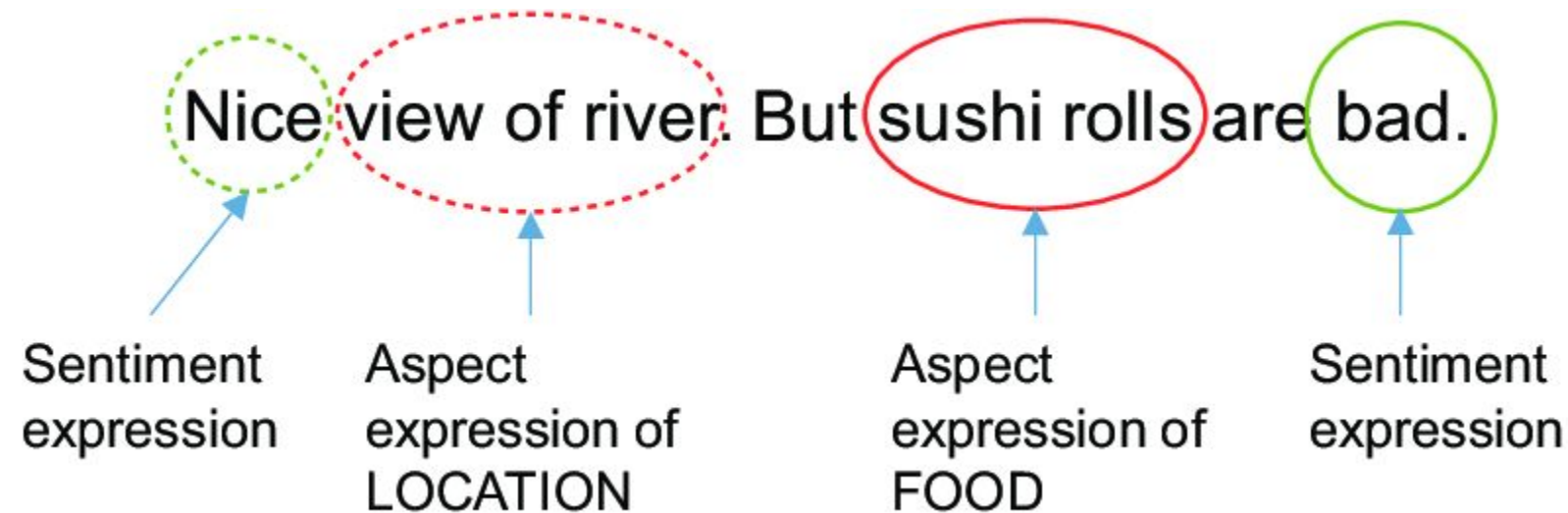
Sentiment Analysis

- Simplest task:
 - Is the attitude of this text positive or negative?
- More complex:
 - Rank the attitude of this text from 1 to 5
- Advanced:
 - Detect the target, source, or complex attitude types

Aspect Based Sentiment Analysis



ASTE vs ASQP



Aspect Sentiment Triplet Extraction:

Triplets of (*Aspect*, *Opinion*, *Sentiment*)

1. (view of river, nice, POS)
2. (sushi rolls, bad, NEG)

Aspect Sentiment Quad Prediction:
(Also known as ACOS)

Quadruplets of (*Asp*, *Category*, *Opi*, *Sent*)

1. (view of river, LOCATION, nice, POS)
2. (sushi rolls, FOOD, bad, NEG)

Sentiment Analysis

A Baseline Algorithm

Sentiment Classification in Movie Reviews

- Polarity detection:
 - Is an IMDB movie review positive or negative?
- Data: *Polarity Data 2.0*:
 - <http://www.cs.cornell.edu/people/pabo/movie--review--data>

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP--2002, 79—86.

Bo Pang and Lillian Lee. 2004. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. ACL, 271--278

IMDB data in the Pang and Lee database



when _star wars_ came out some twenty years ago , the image of traveling throughout the stars has become a commonplace image . [...]

when han solo goes light speed , the stars change to bright lines , going towards the viewer in lines that converge at an invisible point .

cool .

october sky offers a much simpler image—that of a single white dot , traveling horizontally across the night sky . [...]



“ snake eyes ” is the most aggravating kind of movie : the kind that shows so much potential then becomes unbelievably disappointing .

it’s not just because this is a brian depalma film , and since he’s a great director and one who’s films are always greeted with at least some fanfare .

and it’s not even because this was a film starring nicolas cage and since he gives a brauvara performance , this film is hardly worth his talents .

Baseline Algorithm

- Tokenization
- Feature Extraction
- Classification using different classifiers
 - Classical Machine learning (e.g. Naïve Bayes, SVM) – Use scikit-learn library
 - Deep Learning Methods (CNN, RNN, Transformers) – Use pytorch and transformers libraries.

Sentiment Tokenization Issues

- Deal with HTML and XML markup
- Twitter mark--up (mentions, hash tags, URLs)
- Capitalization (preserve for words in all caps)
- Phone numbers, dates
- Emoticons
- Can use NLTK Library's Tweet tokenizer

Extracting Features for Classical ML

- How to handle negation
 - I **didn't** like this movie
vs
 - I really like this movie
- Which words to use?
 - Only adjectives
 - All words
 - All words turns out to work better, at least on this data

Negation for classical ML

Das, Sanjiv and Mike Chen. 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In Proceedings of the Asia Pacific Finance Association Annual Conference (APFA).
Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP-2002, 79—86.

Add NOT_ to every word between negation and following punctuation:

didn't like this movie , but I



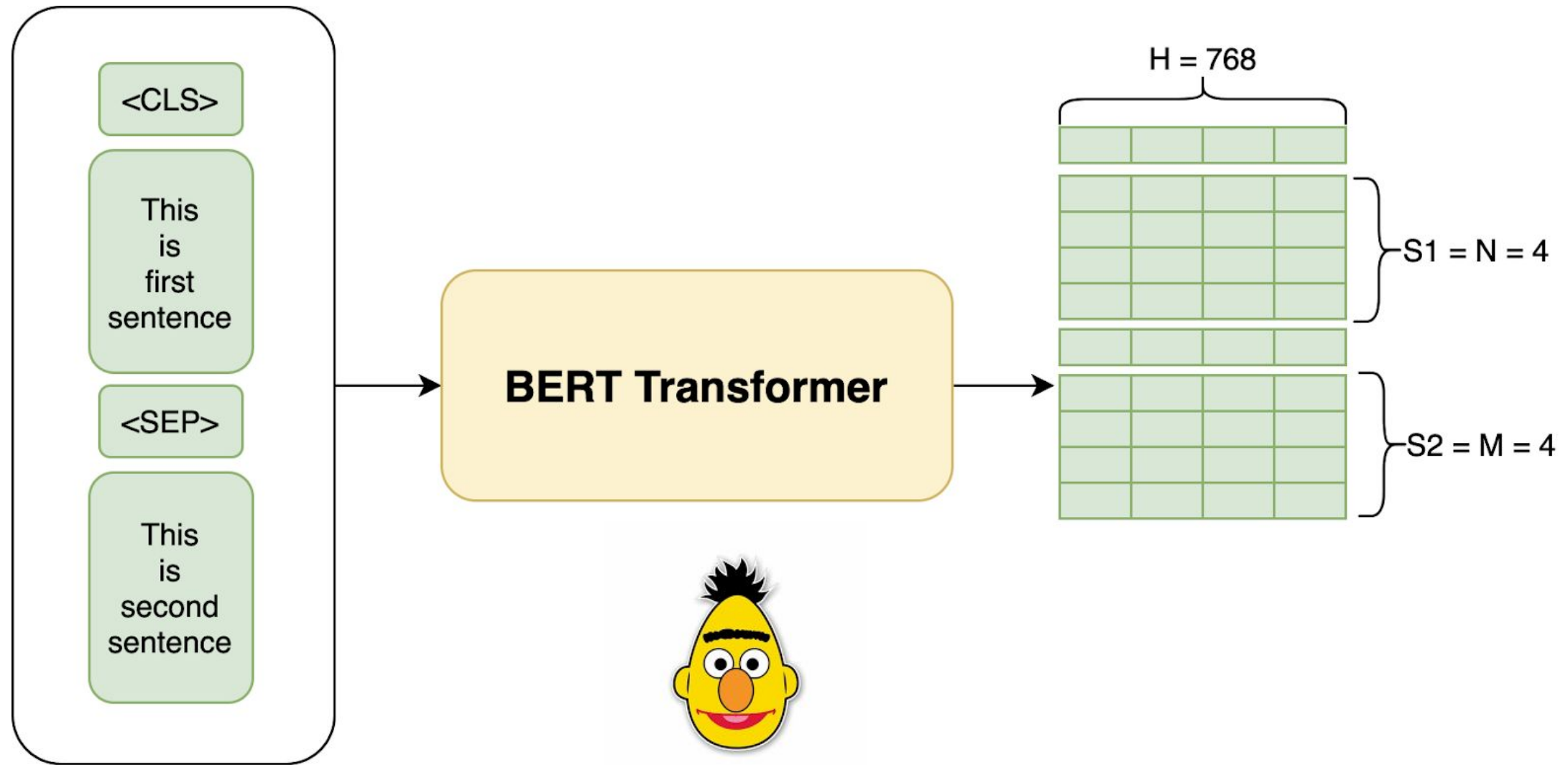
didn't NOT_like NOT_this NOT_movie but I

Binarized (Boolean feature) vs TFIDF

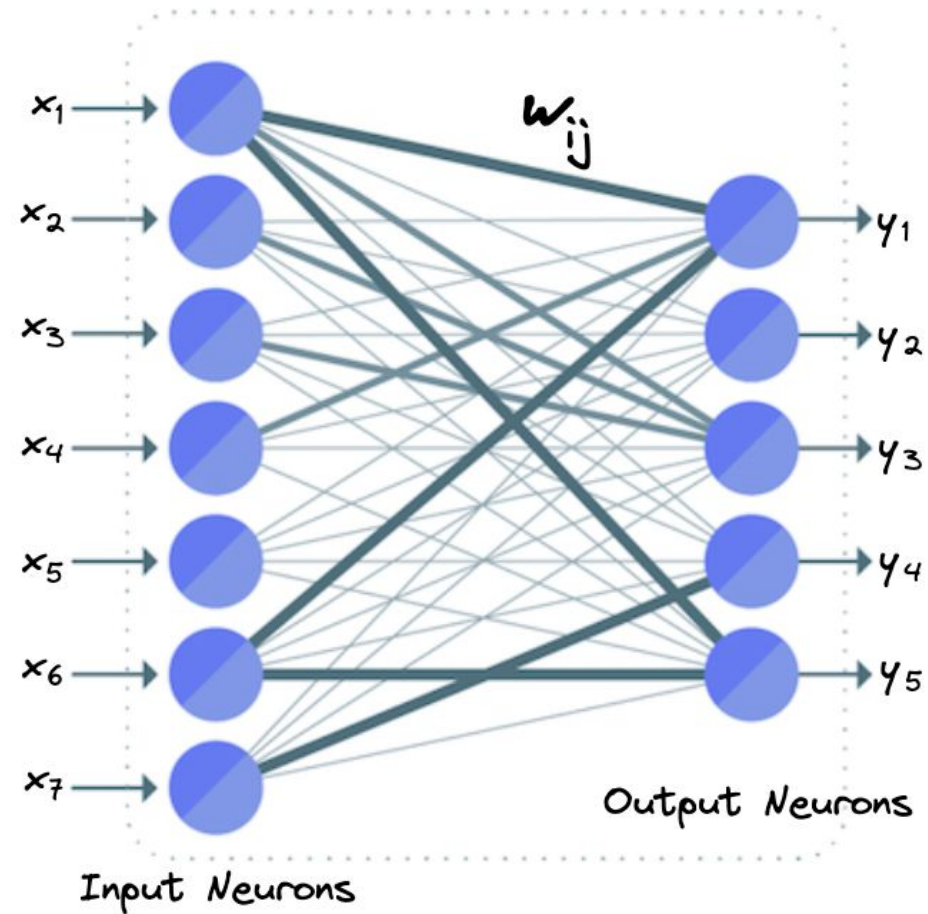
- Intuition:
 - For sentiment (and probably for other text classification domains)
 - Word occurrence may matter more than word frequency
 - The occurrence of the word *fantastic* tells us a lot
 - The fact that it occurs 5 times may not tell us much more
- Boolean Features
 - Clips all the word counts in each document at 1
- Boolean Features
- TFIDF
 - Weighs word importance according to inverse document frequency



Transformers Primer



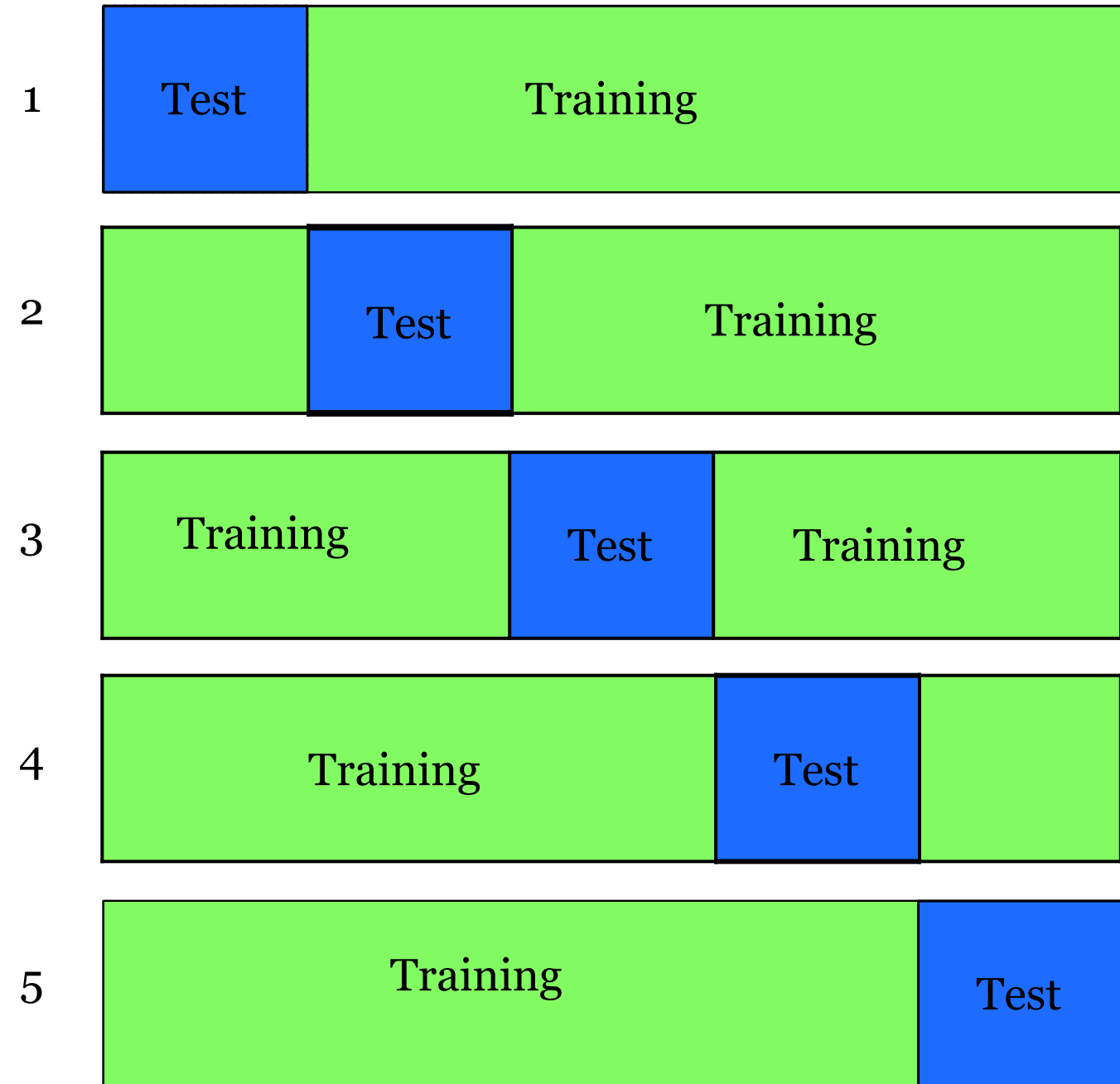
Classification by Fully Connected Layer



Cross—Validation

- Break up data into 10 folds
 - (Equal positive and negative inside each fold?)
- For each fold
 - Choose the fold as a temporary test set
 - Train on 9 folds, compute performance on the test fold
- Report average performance of the 10 runs

Iteration



Sentiment Analysis

Challenges

What makes reviews hard to classify?

- Subtlety:
 - Perfume review in *Perfumes: the Guide*:
 - “If you are reading this because it is your darling fragrance, please wear it at home exclusively, and tape the windows shut.”
 - Dorothy Parker on Katherine Hepburn
 - “She runs the gamut of emotions from A to B”

Explicit vs Implicit



Explicit: Great **food** but the **service** is dreadful



Implicit: The **waiter** poured **water** on my hand and walked away



Thwarted Expectations and Ordering Effects

- “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**.”
- Well as usual Keanu Reeves is nothing special, but surprisingly, the **very talented** Laurence Fishbourne is **not so good** either, I was surprised.

Thank You!

Resources for DL:

- <https://pytorch.org/tutorials/beginner/basics/intro.html>
- <https://huggingface.co/learn/nlp-course/chapter1/1>