Language Understanding

13 - Sentiment Analysis and Opinion Mining

Hossein Zeinali

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

1 star	2	3	4 stars	5 stars

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."

Bing Shopping

HP Officejet 6500A E710N Multifunction Printer

Product summary | Find best price | **Customer reviews** | Specifications | Related items

$121.53 - $242.39 (14 stores)

☐ Compare

Average rating ★★★½☆ (144)

★★★★★		(55)
★★★★☆		(54)
★★★☆☆		(10)
★★☆☆☆		(6)
★☆☆☆☆		(23)
☆☆☆☆☆		(0)

Most mentioned

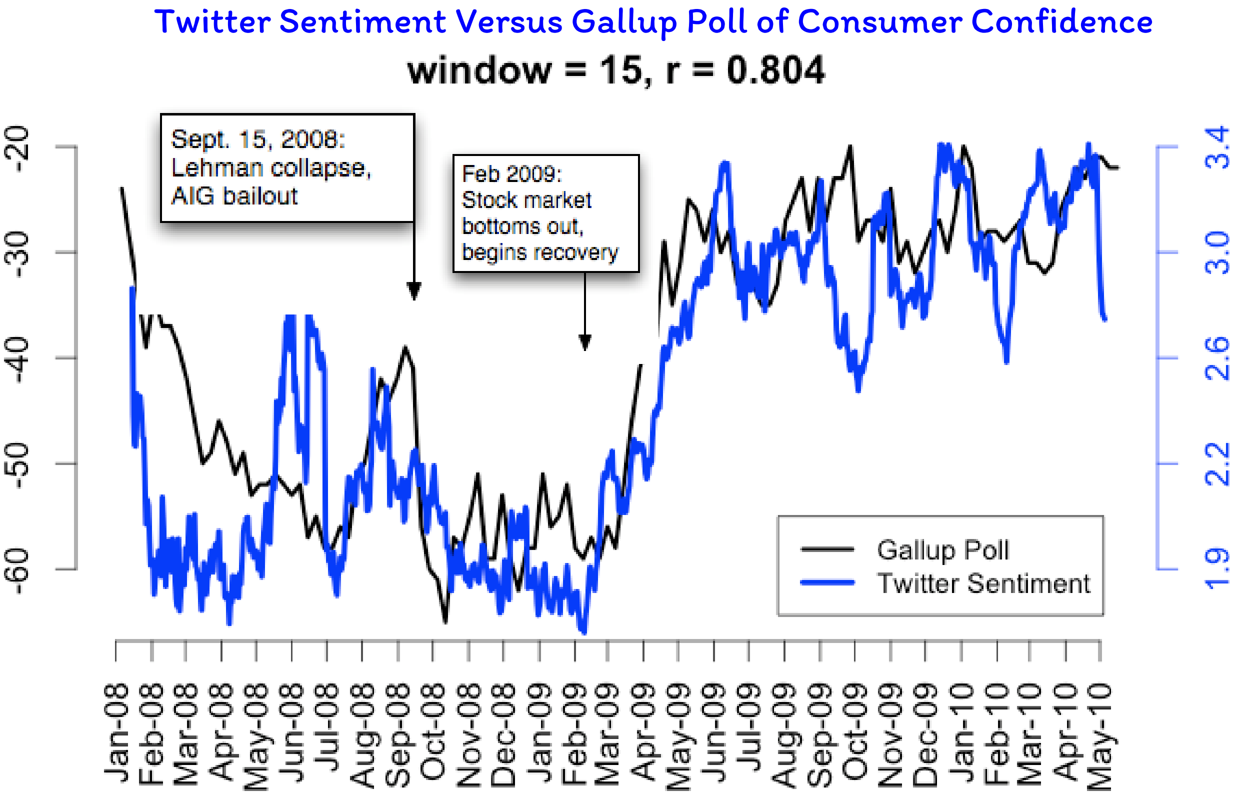
Performance	(57)
Ease of Use	(43)
Print Speed	(39)
Connectivity	(31)

More ▾

Show reviews by source

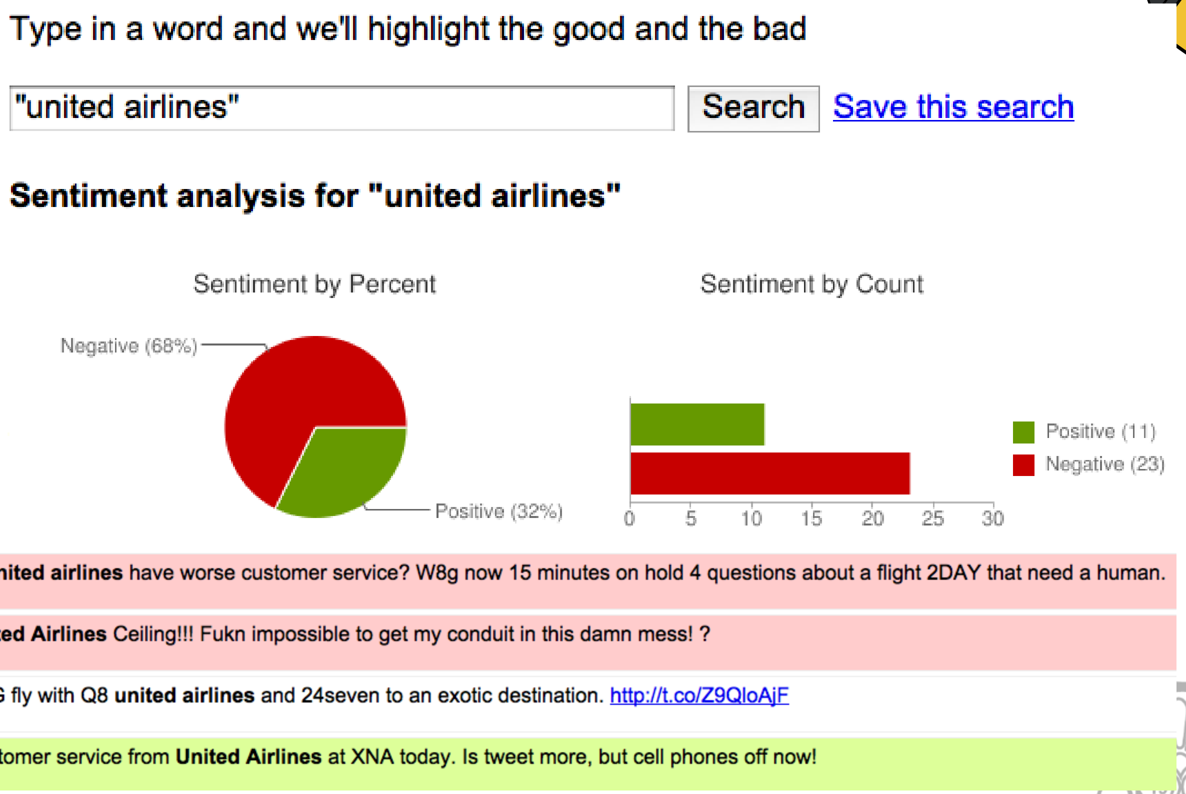
- Best Buy (140)
- CNET (5)
- Amazon.com (3)

Measuring Consumer Confidence



Target Sentiment on Twitter

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



Other Names of Sentiment Analysis

- 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**: evaluation of a major event / brief kind of sentiment
  - *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood**: diffuse affect state, 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**
  - **Holder (source)** of attitude
  - **Target (aspect)** of attitude
  - **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*
  - **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

Online User Reviews

- Heavily influence customer decisions:
  - Travel booking
  - Box Office success
  - Shopping
- Incredibly rich data source:
  - 6.3 million Yelp reviews written in 2010
  - 27.3 million in 2017

Document-level Sentiment Analysis

Rating: ★★

I had a very mixed experience at The Stand. The burger and fries were good. The chocolate shake was divine! The drive-thru was horrible. It took us at least 30 minutes to order. We complained about the wait and got no apology. I would go back because the food is good, but my only hesitation is the wait.

[insert favorite neural net here]

↓

Predicted rating: **

Fine-grained Sentiment Analysis

Rating: ★★

I had a very mixed experience at The Stand. The burger and fries were good. The chocolate shake was divine! The drive-thru was horrible. It took us at least 30 minutes to order. We complained about the wait and got no apology. I would go back because the food is good, but my only hesitation is the wait.

Positive:

- The burger and fries were good.
- The chocolate shake was divine!
- I would go back because the food is good.

Negative:

- The drive-thru was horrible.
- It took us at least 30 minutes to order.
- We complained about the wait and got no apology.
- My only hesitation is the wait.

Polarity-based Opinion Extraction

Rating: ★★

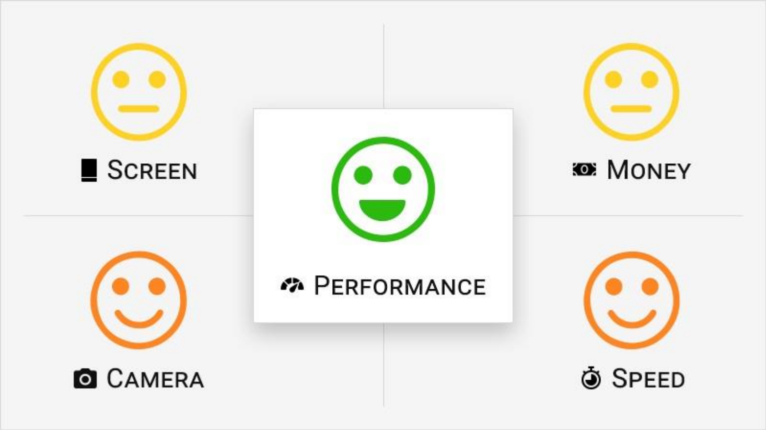
I had a very mixed experience at The Stand. The burger and fries were good. The chocolate shake was divine! The drive-thru was horrible. It took us at least 30 minutes to order. We complained about the wait and got no apology. I would go back because the food is good, but my only hesitation is the wait.

Very negative ↔ Very positive

[+1.00]	**The chocolate shake was divine**
[+0.86]	**I would go back because the food is good**
[+0.50]	**The burger and fries were good**
[−0.05]	I had a very mixed experience at The Stand.
[−0.10]	but my only hesitation is the wait
[−0.10]	and got no apology
[−0.25]	We complained about the wait
[−0.43]	**It took us at least 30 minutes to order**
[−0.89]	**The drive-thru was horrible**

Aspect Based Sentiment Analysis

- Is a technique that takes into consideration the terms related to the aspects and identifies the sentiment associated with each aspect.
- ABSA model requires *aspect categories* and its corresponding *aspect terms*



Sentiment Datasets

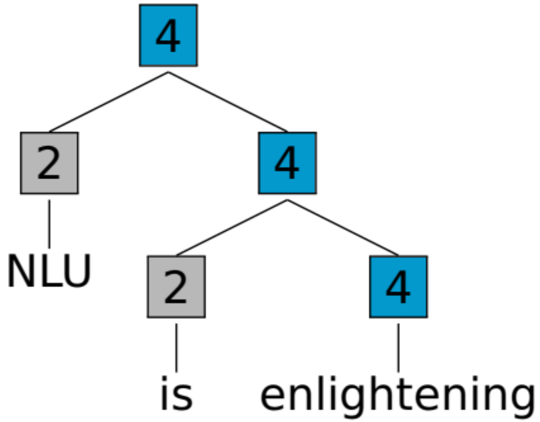
IMDb

- The IMDb dataset is a binary sentiment analysis dataset
- Consisting of 50,000 reviews from the Internet Movie Database (IMDb)
- Labeled as positive or negative
- Only highly polarizing reviews are considered
  - A negative review has a score ≤ 4 out of 10, and a positive review has a score ≥ 7 out of 10.
- No more than 30 reviews are included per movie.
- Models are evaluated based on accuracy.

Stanford Sentiment Treebank (SST)

- Contains 215,154 phrases with fine-grained sentiment labels in the parse trees of 11,855 sentences in movie reviews.
- Models are evaluated either on fine-grained (SST-1) or binary classification (SST-2) based on accuracy.

Label	Meaning
0	very negative
1	negative
2	neutral
3	positive
4	very positive

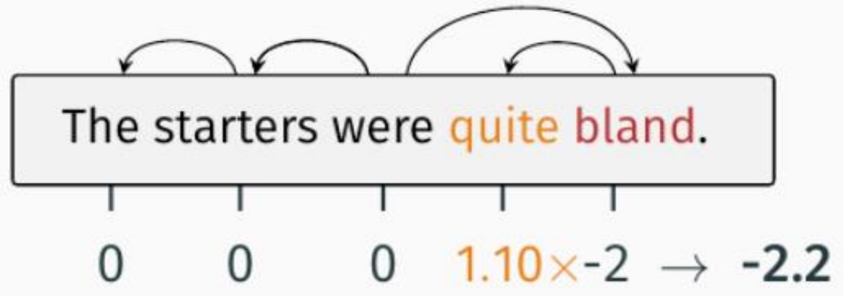


Yelp

- The Yelp Review dataset consists of more than 500,000 Yelp reviews.
    - User Reviews and Recommendations of Best Restaurants, Shopping, Nightlife, Food, Entertainment, Things to Do, and Services at *Yelp*.
- There is both a binary and a fine-grained (five-class) version of the dataset.
- Models are evaluated based on error (1 - accuracy; lower is better).

Sentiment Analysis Methods

Fine-grained Sentiment: Unsupervised

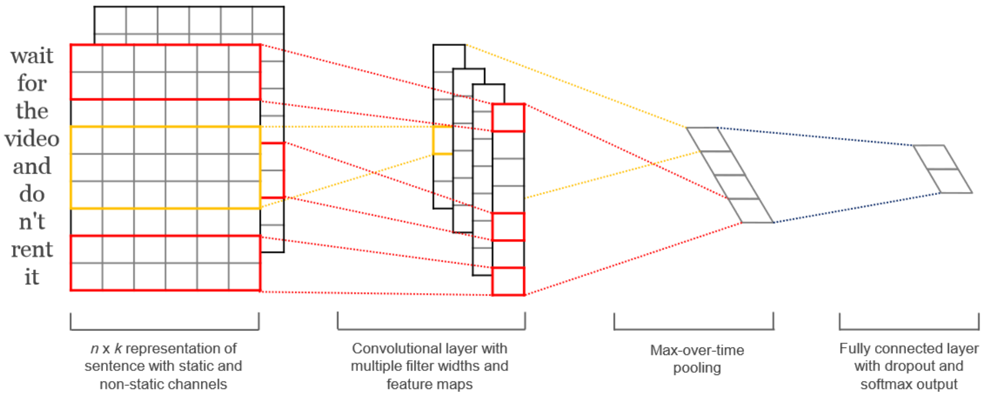


SO-CAL: Semantic Orientation CALculator
Taboada et al., Lexicon-Based Methods for Sentiment Analysis, 2011

Adjective:		**Intensifier:**	
disgusting	-5	slightly	0.50
terrible	-4	somewhat	0.70
bland	-2	pretty	0.90
so-so	-1	quite	1.10
okay	1	really	1.15
great	2	very	1.25
amazing	4	extraordinarily	1.50
divine	5	(the) most	2.00

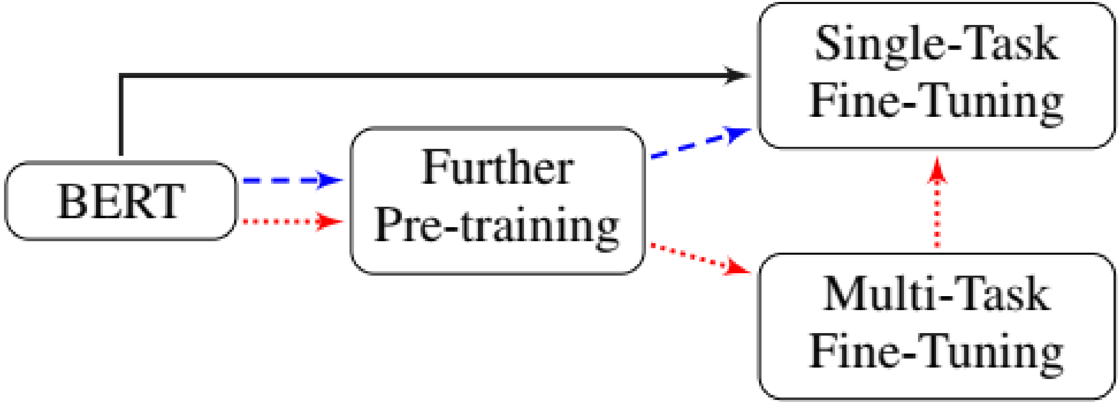
Fine-grained Sentiment Using CNN

- Segment CNN
- Multiple convolutional filters of varying length
- Max-over-time pooling
- Successful for sentence classification
- Segment encoder in larger networks
- Requires expensive annotations



Source: Convolutional Neural Networks for Sentence Classification 2014

Fine-Tune BERT for Text Classification



- **Further Pre-training:**
  - Further pre-train BERT with target domain data instead of the general domain.

Source: How to Fine-Tune BERT for Text Classification? 2020

Fine-Tune BERT for Text Classification

Domain	sentiment		
Dataset	IMDb	Yelp P.	Yelp F.
IMDb	**4.37**	2.18	29.60
Yelp P.	5.24	1.92	29.37
Yelp F.	5.18	1.94	29.42
all sentiment	4.88	**1.87**	29.25

Performance of in-domain and cross-domain further pre-training

Layer	Test error rates(%)
Layer-0	11.07
Layer-1	9.81
Layer-2	9.29
Layer-3	8.66
Layer-4	7.83
Layer-5	6.83
Layer-6	6.83
Layer-7	6.41
Layer-8	6.04
Layer-9	5.70
Layer-10	5.46
Layer-11	**5.42**
First 4 Layers + concat	8.69
First 4 Layers + mean	9.09
First 4 Layers + max	8.76
Last 4 Layers + concat	5.43
Last 4 Layers + mean	5.44
Last 4 Layers + max	**5.42**
All 12 Layers + concat	5.44

T5 Model Results on SST-2

Model	SST-2 Accuracy
Previous best	97.1[a]
T5-Small	91.8
T5-Base	95.2
T5-Large	96.3
T5-3B	97.4
T5-11B	**97.5**

Best Results on SST-2

RANK	MODEL	ACCURACY	PAPER	YEAR
1	T5-3B	97.4	Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer	2019
2	ALBERT	97.1	ALBERT: A Lite BERT for Self-supervised Learning of Language Representations	2019
3	T5-11B	97.1	Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer	2019
4	XLNet (single model)	97	XLNet: Generalized Autoregressive Pretraining for Language Understanding	2019
5	RoBERTa	96.7	RoBERTa: A Robustly Optimized BERT Pretraining Approach	2019
6	FLOATER-large	96.7	Learning to Encode Position for Transformer with Continuous Dynamical Model	2020
7	T5-Large	96.3	Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer	2019
8	MT-DNN	95.6	Multi-Task Deep Neural Networks for Natural Language Understanding	2019
9	T5-Base	95.2	Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer	2019

Source: https://paperswithcode.com

Best Results on Yelp Binary

RANK	MODEL	ERROR	PAPER	YEAR
1	BERT large	1.89	Unsupervised Data Augmentation for Consistency Training	2019
2	BERT large finetune UDA	2.05	Unsupervised Data Augmentation for Consistency Training	2019
3	ULMFiT	2.16	Universal Language Model Fine-tuning for Text Classification	2018
4	DPCNN	2.64	Deep Pyramid Convolutional Neural Networks for Text Categorization	2017
5	DRNN	2.73	Disconnected Recurrent Neural Networks for Text Categorization	2018
6	CNN	2.90	Supervised and Semi-Supervised Text Categorization using LSTM for Region Embeddings	2016
7	Block-sparse LSTM	3.27	GPU Kernels for Block-Sparse Weights	2017
8	CCCapsNet	3.52	Compositional coding capsule network with k-means routing for text classification	2018

Source: https://paperswithcode.com

Thanks for your attention

References and IP Notice

- Some slides from Daniel Jurafsky's slides on sentiment analysis.
- Some slides from Mirella Lapata's slides on sentiment analysis.
- The mentioned papers for each methods.
- Some graphics from Slidesgo online template.