Sentiment Analysis

Annotated bibliography

- 6 citations versus 12 citations

Programming Assignments

Question with spreadsheet

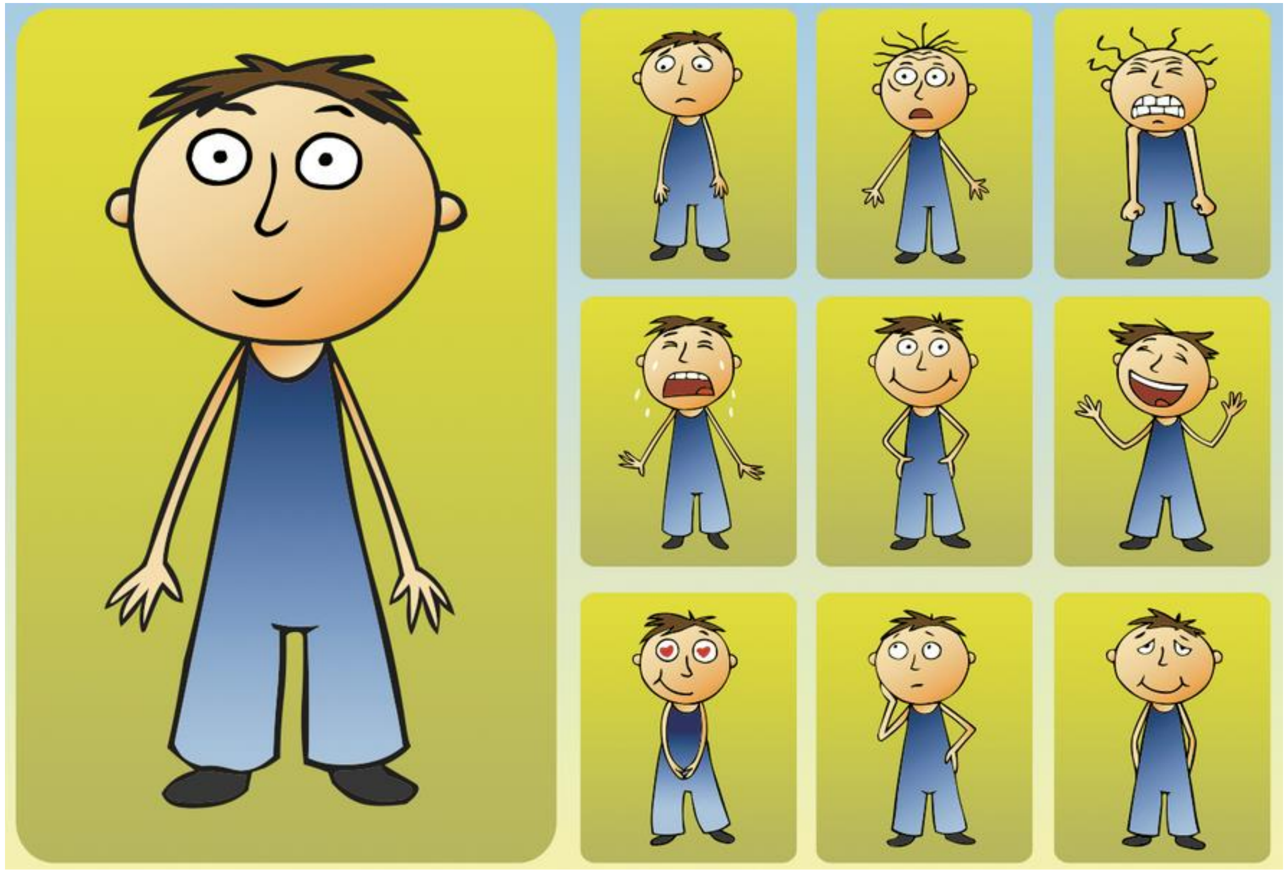
Hit record

What is sentiment?

- Sentiment = feelings
  - attitudes
  - emotions
  - opinions

- Subjective impressions

  Not facts!



What is sentiment?

- Determining if someone is:
  - For/against
  - Like/dislike
  - Good/bad

- For example:
  - Polarity:
    - state of two opposite opinions

Semantic orientation

opinion on a feature f states whether the opinion is
positive, negative or neutral
with sentiment in between

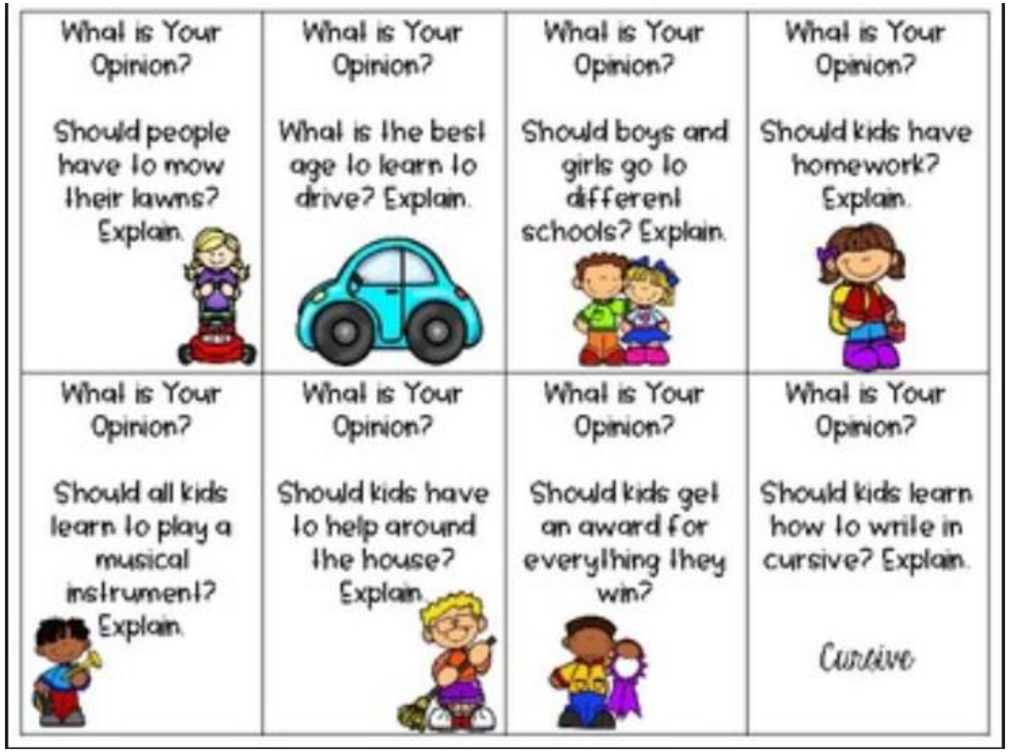
What factors makes one more difficult to classify than the other?

NLP tasks that use semantic analysis

- Information extraction
  - discarding subjective information from the returned results

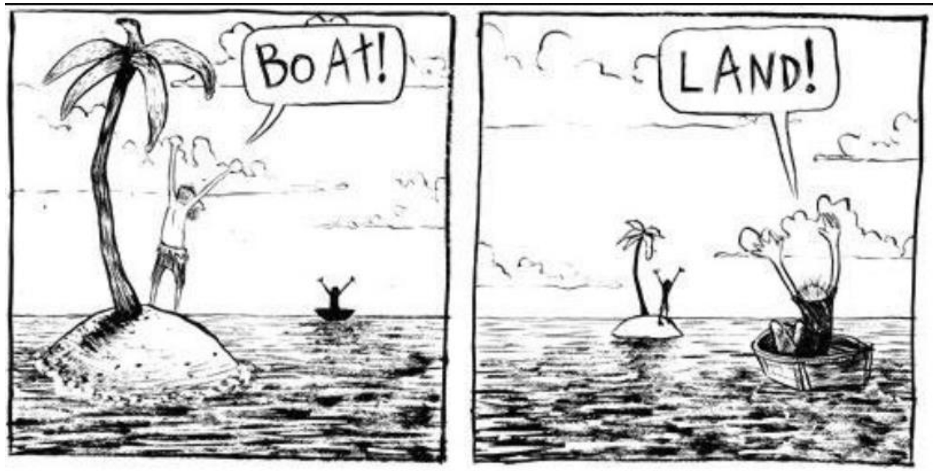
NLP tasks that use semantic analysis

- Information extraction
  - discarding subjective information from the returned results
- Question answering
  - recognizing opinion-oriented questions



NLP tasks that use semantic analysis

- Information extraction
  - discarding subjective information from the returned results
- Question answering
  - recognizing opinion-oriented questions
- Summarization
  - accounting for multiple points of view and summarizing each of them appropriately



NLP tasks that use semantic analysis

- Information extraction
  - discarding subjective information from the returned results
- Question answering
  - recognizing opinion-oriented questions
- Summarization
  - accounting for multiple points of view and summarizing each of them appropriately
- Bias detection
  - in news sources

Sentiment Analysis as a Classification Problem

- Comparatively few categories compared to other text categorization tasks
  - negative/positive
  - three star

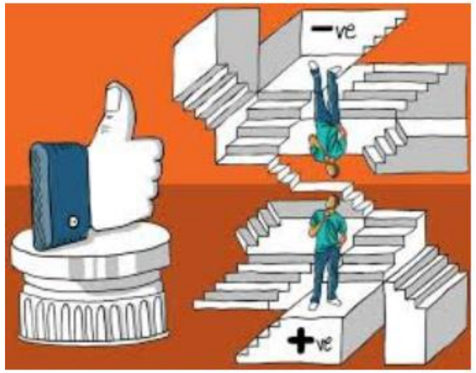
Sentiment Analysis as a Classification Problem

- Comparatively few categories compared to other text categorization tasks
  - negative/positive
  - three star

- Crosses domains, topics and users

Sentiment Analysis as a Classification Problem

- Comparatively few categories compared to other text categorization tasks
  - negative/positive
  - three star



- Crosses domains, topics and users

- Categories are not independent

Sentiment Analysis as a Classification Problem

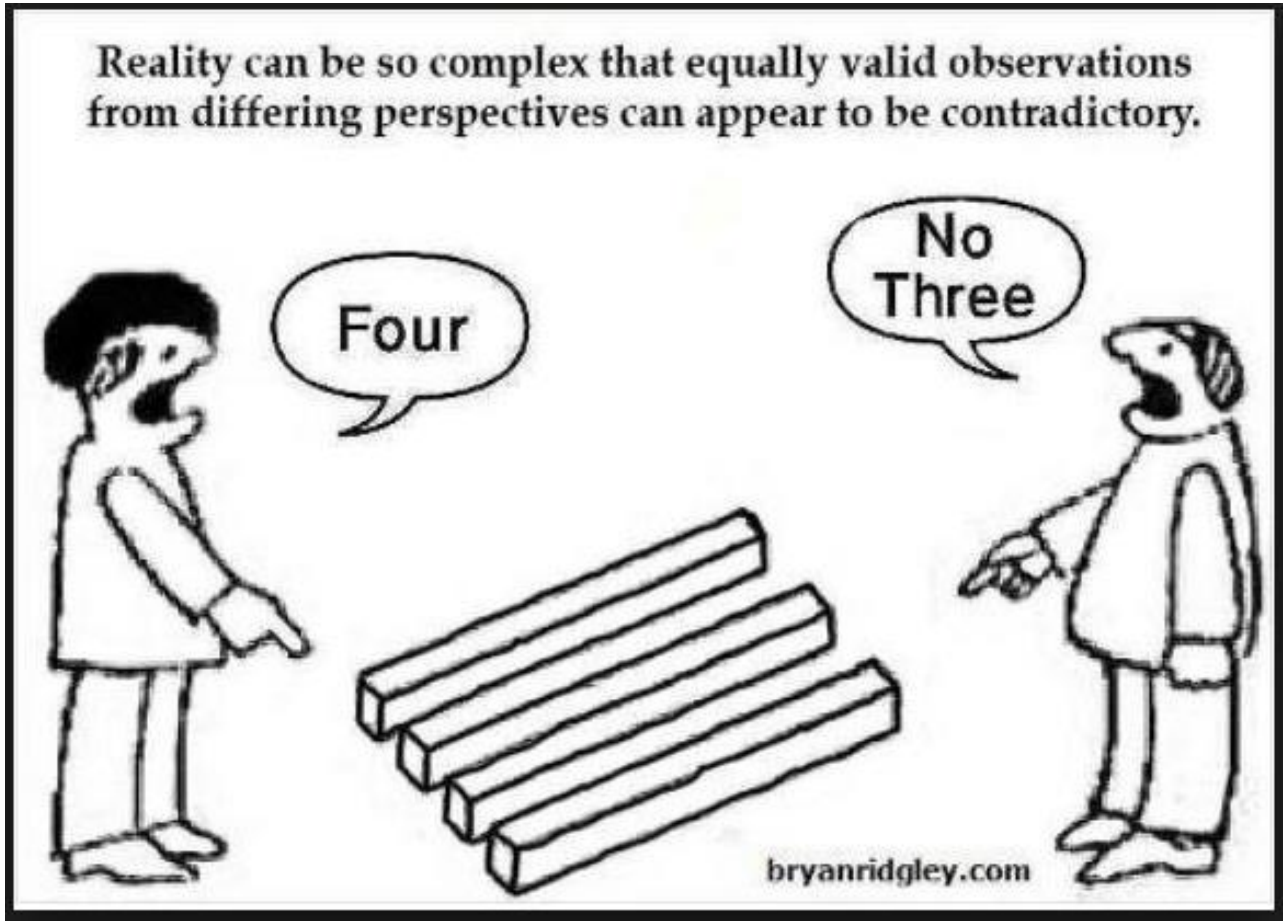
- Comparatively few categories compared to other text categorization tasks
  - negative/positive
  - three star

- Crosses domains, topics and users

- Categories are not independent

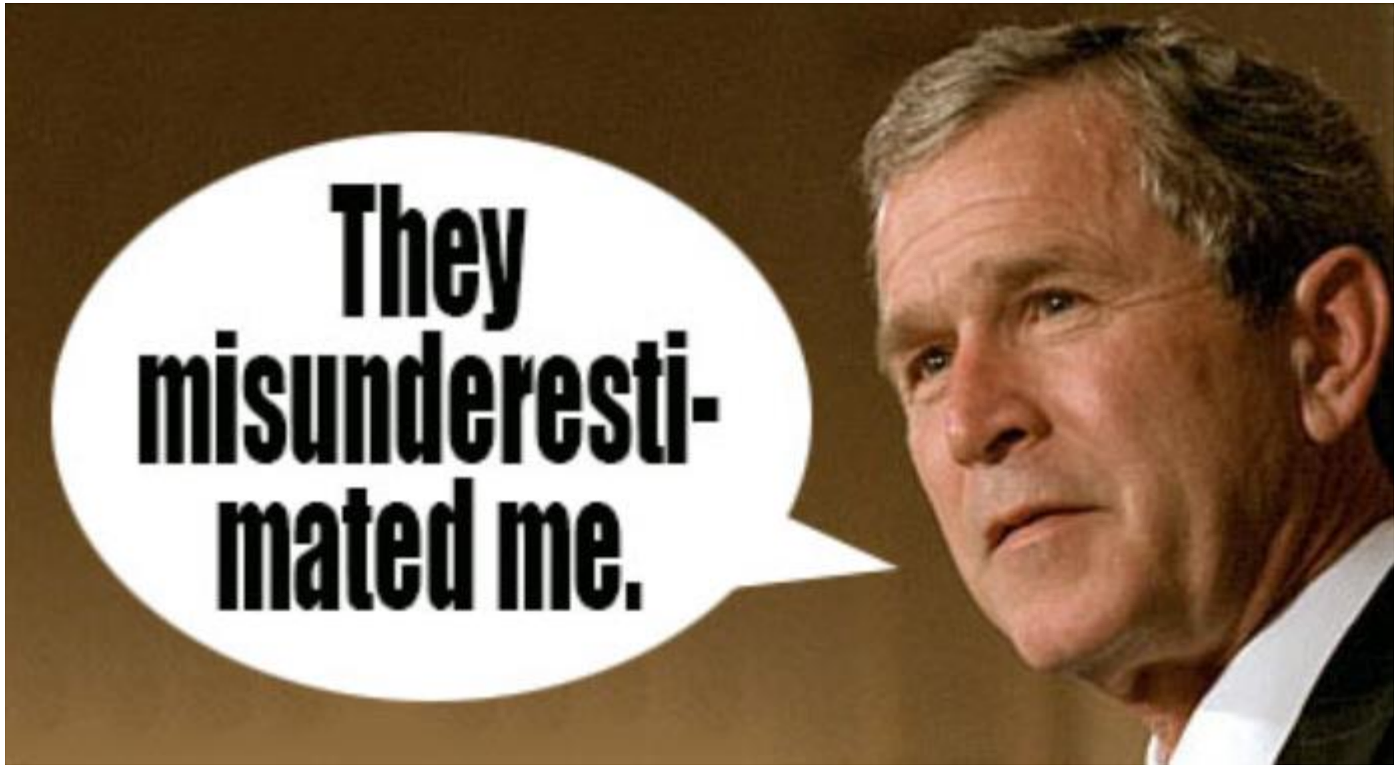
- Characteristics of answers to opinion-based questions are different from fact-based questions

Challenges in semantic analysis

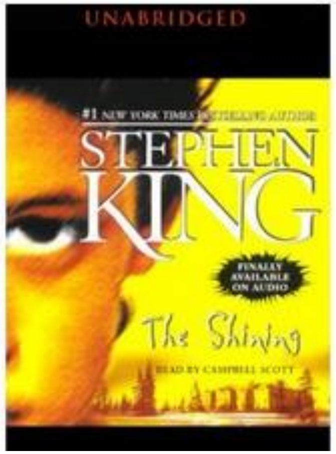


Challenges in semantic analysis

- People express opinions in complex ways



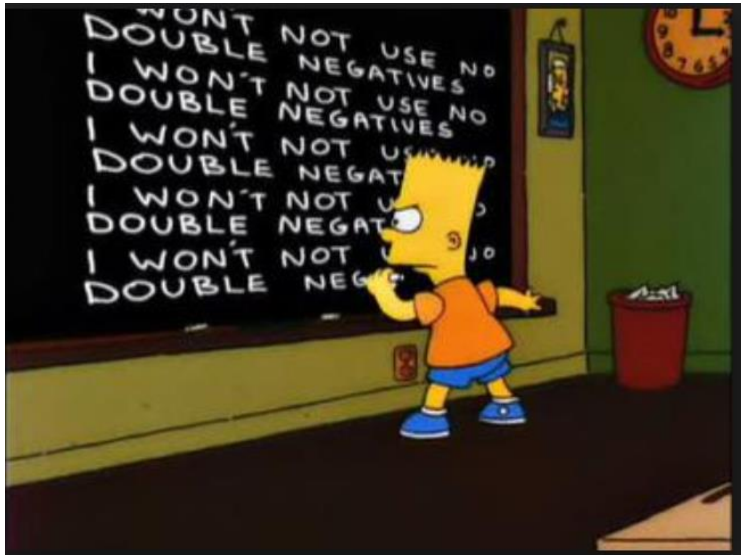
Challenges in semantic analysis



- People express opinions in complex ways
- In opinion texts, lexical content alone can be misleading

The characters are so real and handled so carefully, that being trapped inside the Overlook is no longer just a freaky experience. You run along with them, filled with dread, from all the horrible personifications of evil inside the hotel's awful walls. There were several times where I actually dropped the book and was too scared to pick it back up. Intellectually, you know it's not real. It's just a bunch of letters and words grouped together on pages. Still, whenever I go into the bathroom late at night, I have to pull back the shower curtain just to make sure.

Challenges in semantic analysis



- People express opinions in complex ways
- In opinion texts, lexical content alone can be misleading
- Intra-textual and sub-sentential reversals, negation, topic change are common

Challenges in semantic analysis

- People express opinions in complex ways
- In opinion texts, lexical content alone can be misleading
- Intra-textual and sub-sentential reversals, negation, topic change are common
- Rhetorical devices such as sarcasm and irony are often imployed

Sarcasm Detection

A letter to a hardware store

Dear <store>,

Yesterday I had the occasion to visit <your competitor>. They had an excellent selection, friendly and helpful sales people, and the lowest prices in town.

You guys suck.

Sincerely,

Jane Doe

Sarcasm Detection

- Strap in for the Hobbit, the only movie trilogy that takes longer to watch than read.

Sarcasm Detection

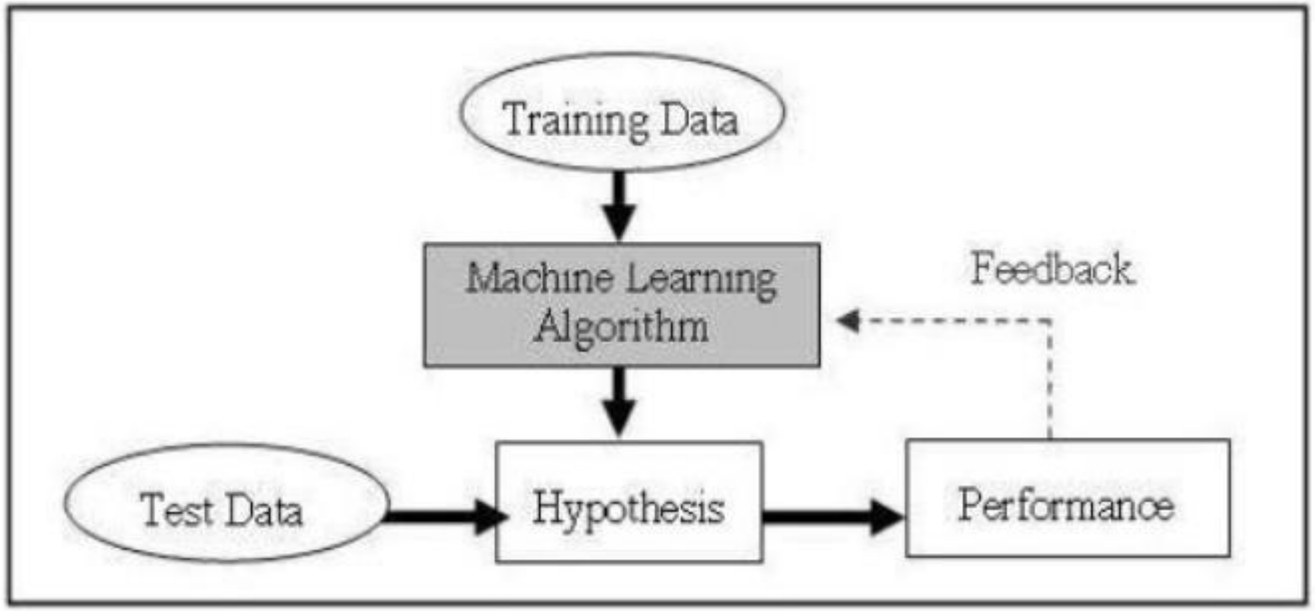
- Strap in for the Hobbit, the only movie trilogy that takes longer to watch than read.
- We loved it – a sure guaranteed cure for insomnia.

Sarcasm Detection

- Strap in for the Hobbit, the only movie trilogy that takes longer to watch than read.
- We loved it – a sure guaranteed cure for insomnia.
- The end is the best part… the time you can finally go home

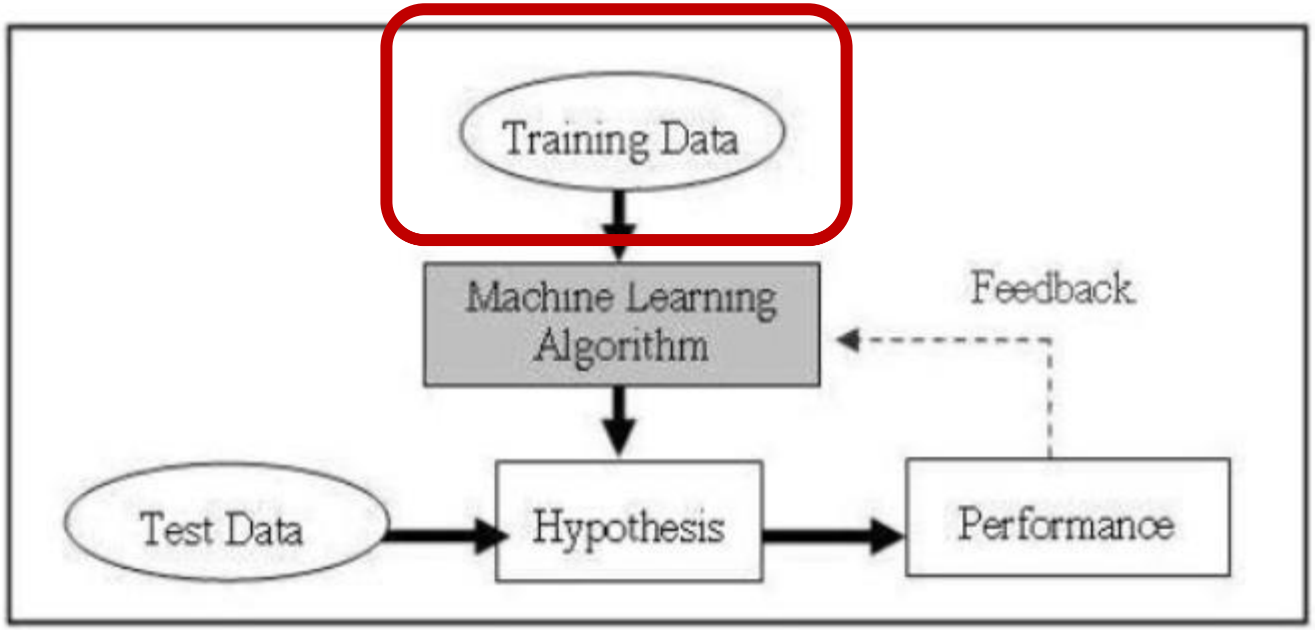
Cast as a classification task

Classification based on supervised learning



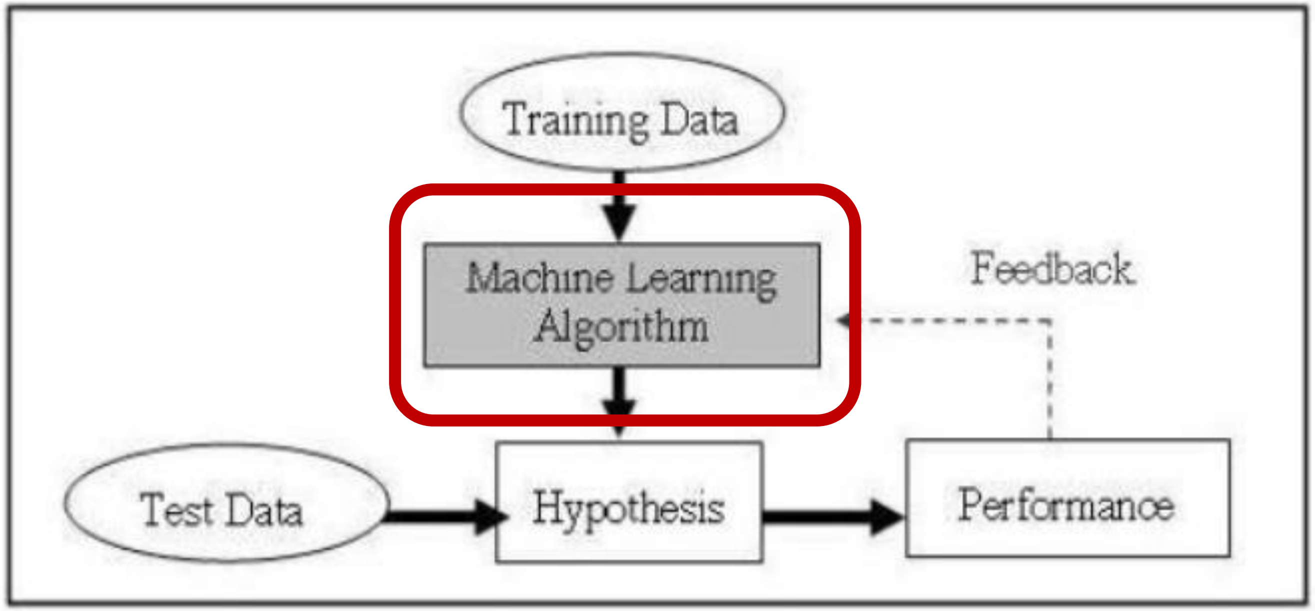
Classification based on supervised learning

- Training Data:
  - manually annotated training data



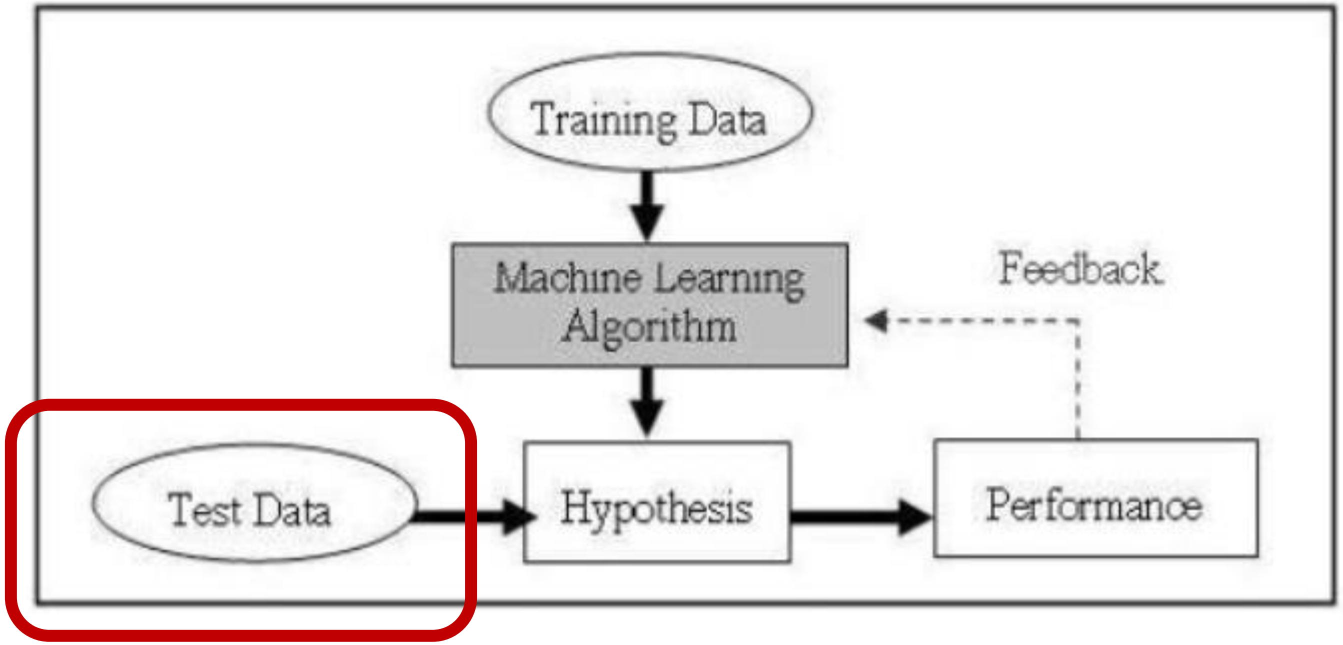
Classification based on supervised learning

- Training Data:
  - manually annotated training data
- Algorithms:
  - Machine learning algorithm
    - Naïve Bayes, SVM, cNN
  - Features
    - how to represent the instances



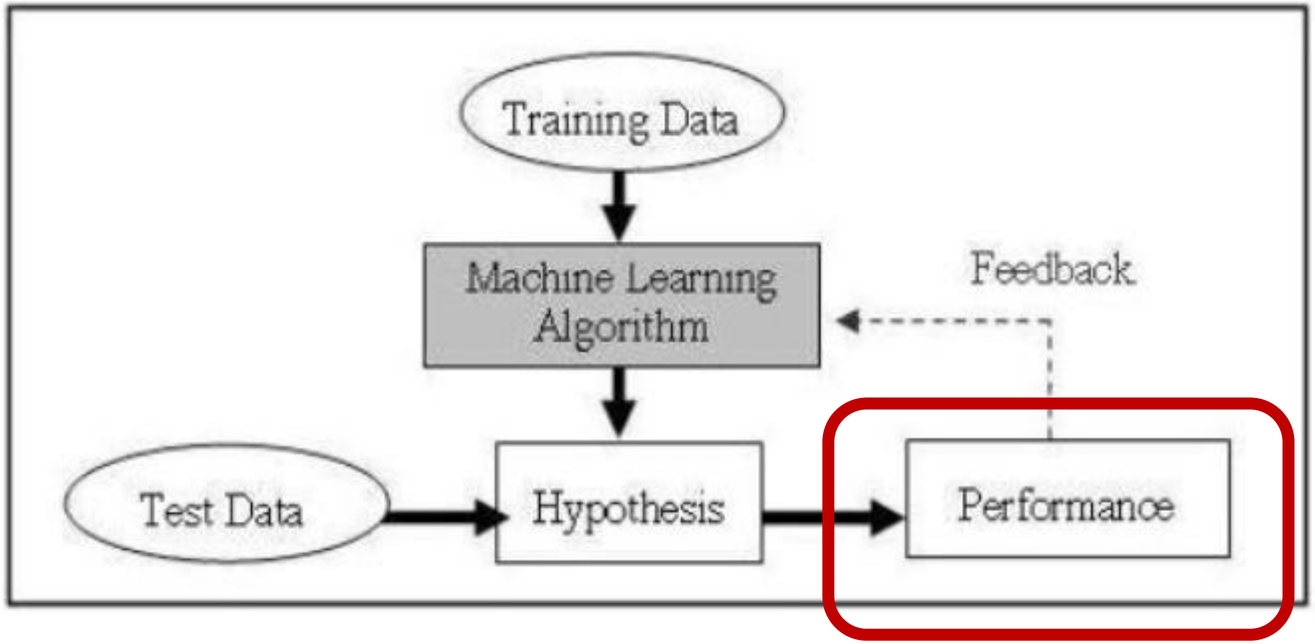
Classification based on supervised learning

- Training Data:
  - manually annotated training data
- Algorithms:
  - Machine learning algorithm
    - Naïve Bayes, SVM, NN
  - Features
    - how to represent the instances
- Test Data
  - manually annotated
  - representative of our training data



Classification based on supervised learning

- Training Data:
  - manually annotated training data
- Algorithms:
  - Machine learning algorithm
    - Naïve Bayes, SVM, NN
  - Features
    - how to represent the instances
- Test Data
  - manually annotated
  - representative of our training data



- Performance
  - how well does our algorithm work
  - Metrics
    - Precision, Recall, F-measure

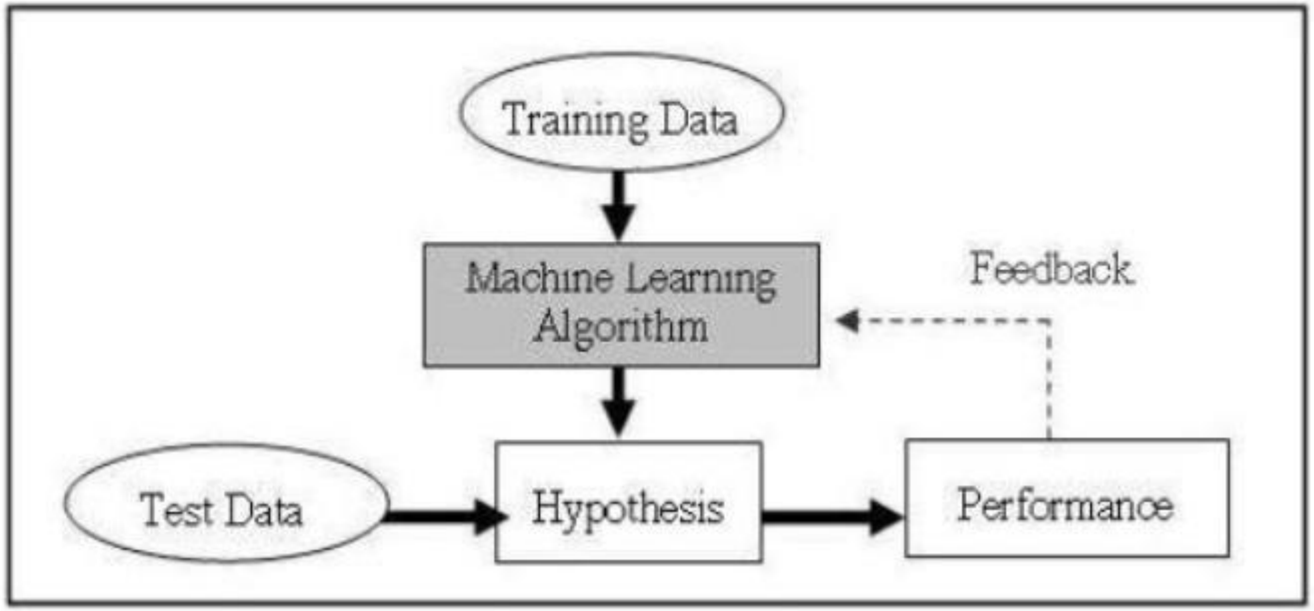
Feature representations

- Feature-based
- Featureless


Feature-based

Pang et al 2002

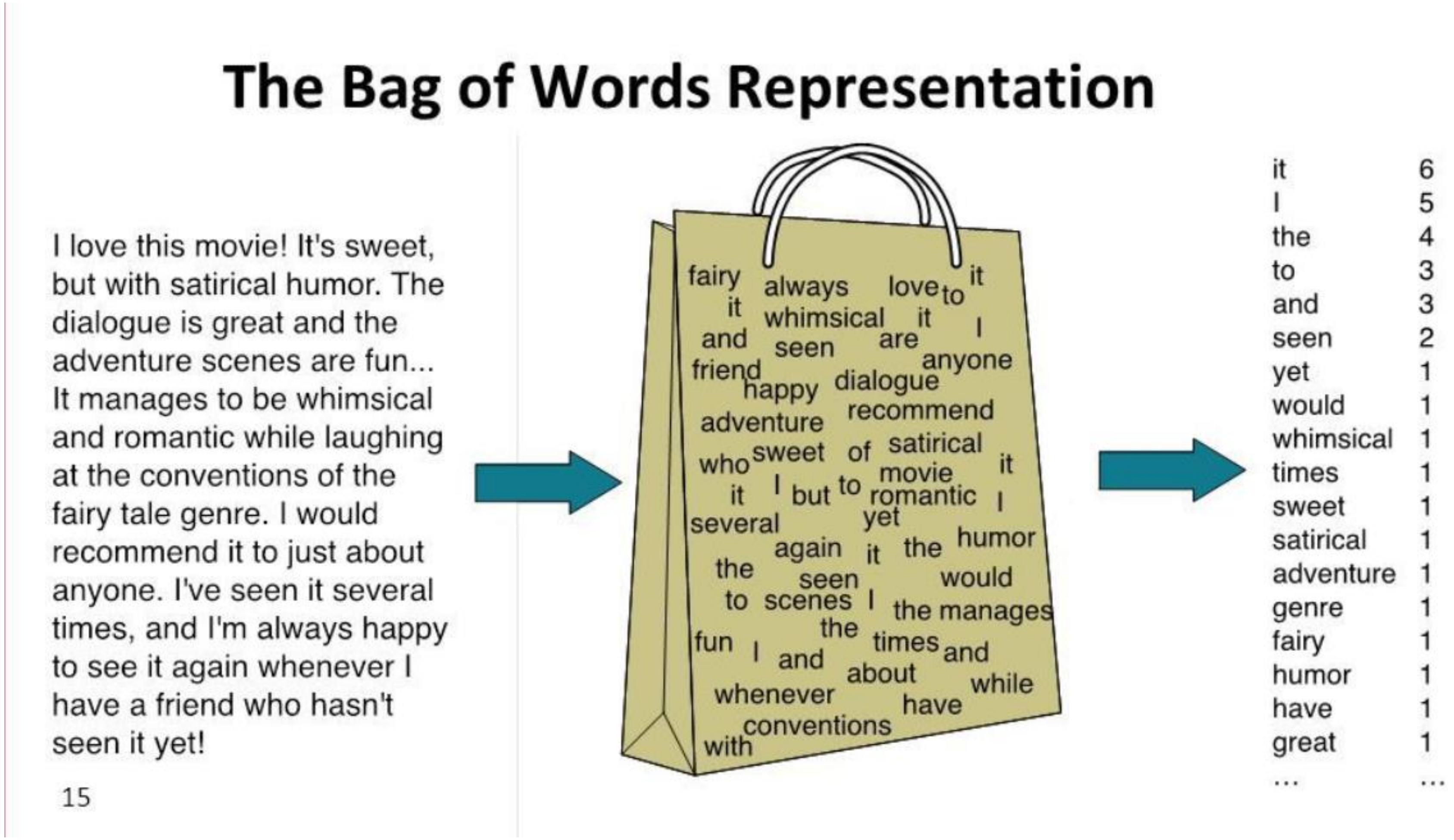
- Training Data: movie reviews
  - 700 positive
  - 700 negative
- Machine Learning
  - Algorithm:
    - Naïve Bayes
    - MaxEnt
    - SVM
  - Features:
    - n-grams
    - POS
- Performance:
  - 3-fold cross validation
  - Accuracy



	Features	# of features	frequency or presence?	NB	ME	SVM
(1)	unigrams	16165	freq.	**78.7**	N/A	72.8
(2)	unigrams	”	pres.	81.0	80.4	**82.9**
(3)	unigrams+bigrams	32330	pres.	80.6	80.8	**82.7**
(4)	bigrams	16165	pres.	77.3	**77.4**	77.1
(5)	unigrams+POS	16695	pres.	81.5	80.4	**81.9**
(6)	adjectives	2633	pres.	77.0	**77.7**	75.1
(7)	top 2633 unigrams	2633	pres.	80.3	81.0	**81.4**
(8)	unigrams+position	22430	pres.	81.0	80.1	**81.6**

One of the first papers to present supervised semantic classification

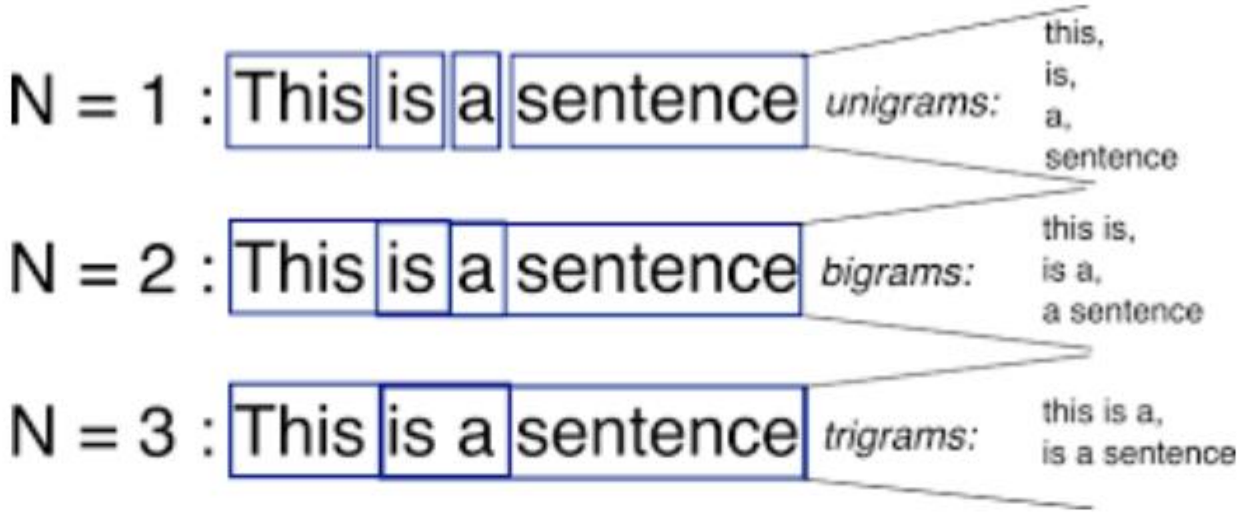
unigrams (aka: bag of words; 1st order co-occurrence vectors)



Subsequent features

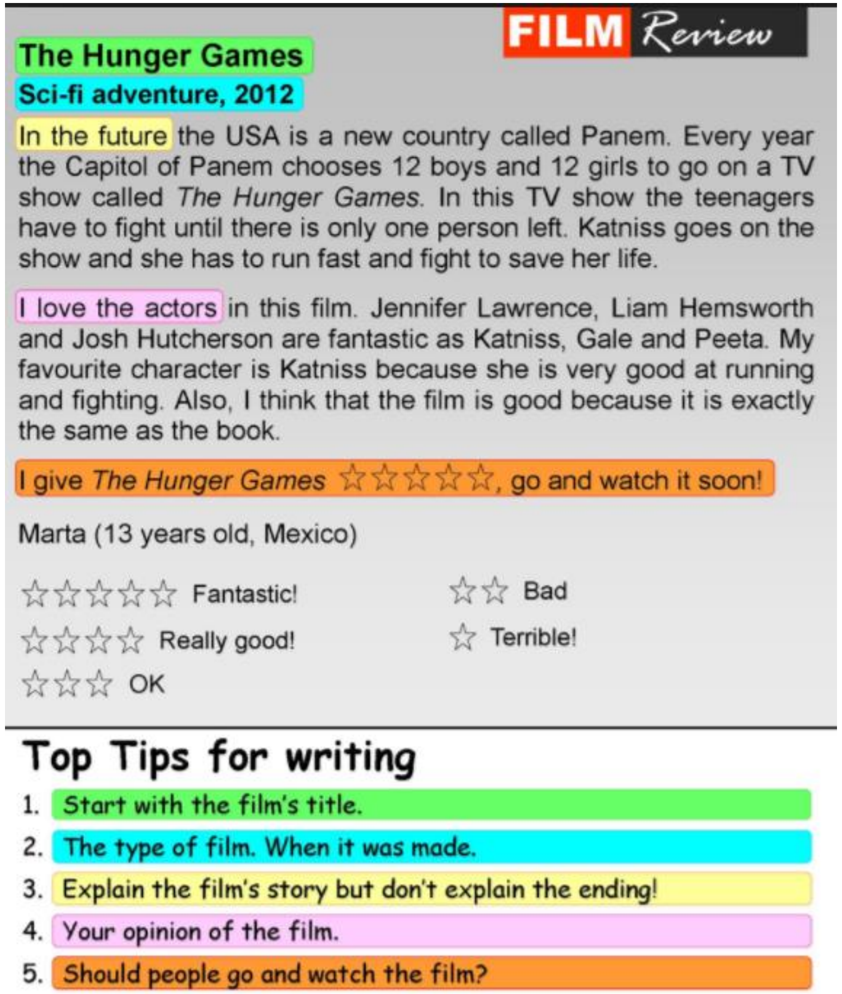
N-grams

- Higher order n-grams
  - Pang et al found unigrams outperform bigrams for movie review classification
  - Dave et al found that bigrams and trigram yield better product-review polarity classification



N-grams

- Higher order n-grams
  - Pang et al found unigrams outperform bigrams for movie review classification
  - Dave et al found that bigrams and trigram yield better product-review polarity classification
- Positional information
  - position of a token within a textual unit
    - middle versus beginning versus end
  - important effect on how much that token affects the overall sentiment



N-grams

- Higher order n-grams
  - Pang et al found unigrams outperform bigrams for movie review classification
  - Dave et al found that bigrams and trigram yield better product-review polarity classification

- Positional information
  - position of a token within a textual unit
    - middle versus beginning versus end
  - important effect on how much that token affects the overall sentiment

- Weighting
  - based on frequency of the term
  - based on the term frequency – inverse document frequency

Acronym expansion

ACRONYM	EXPANSION
SCAM	Samajwadi Party, Congress, Akhilesh, Mayawati
PPP	Punjab, Puducherry, Parivaar
ABCD	Adarsh, Bofors, Coal and Damaad
AK-49	Arvind Kejriwal (and his 49-day first government in Delhi)
RSVP	Rahul, Sonia, Vadra, Priyanka
JAM	Jan Dhan-Aadhaar Mobile
HRIDAY	Heritage Development and Augmentation Yojana
B2B	Bharat to Bhutan
AIM	Atal Innovation Mission
SETU	Self Employment and Talent Utilisation

BUSINESS ACRONYMS

Financial

ACCT	Account
ACR	Accrual
ACV	Actual Cash Value
AGI	Adjusted Gross Income
AGR	Adjusted Gross Revenue
A/R	Accounts Receivable
BS	Balance Sheet
BGT	Budget
COGS	Cost of Goods Sold
CPTAL	Capital
EPS	Earnings Per Share
FIFO	First In, First Out
ROA	Return On Assets

Gen Z Social Media Acronyms And Abbreviations

411	Information
AF	As F----
BAE	Before Anyone Else
BC	Because
FFS	For F---'s Sake
FML	F--- My Life
FWIW	For What It's Worth
HMU	Hit Me Up
IDK	I Don't Know
ILY	I Love You
ISO	In Search Of
JK	Just Kidding
JTM	Just The Messenger
LMAO	Laughing My A-- Off
NVM	Nevermind
NYT	Name Your Trade
Obv	Obviously
OMG	Oh My God
OMW	On My Way
PLS	Please
PSA	Public Service Announcement
RN	Right Now
ROFL	Rolling On The Floor Laughing
SRSLY	Seriously
TMI	Too Much Informatioin
TY	Thank You
WTF	What The F---
YW	You're Welcome

Part-of-speech

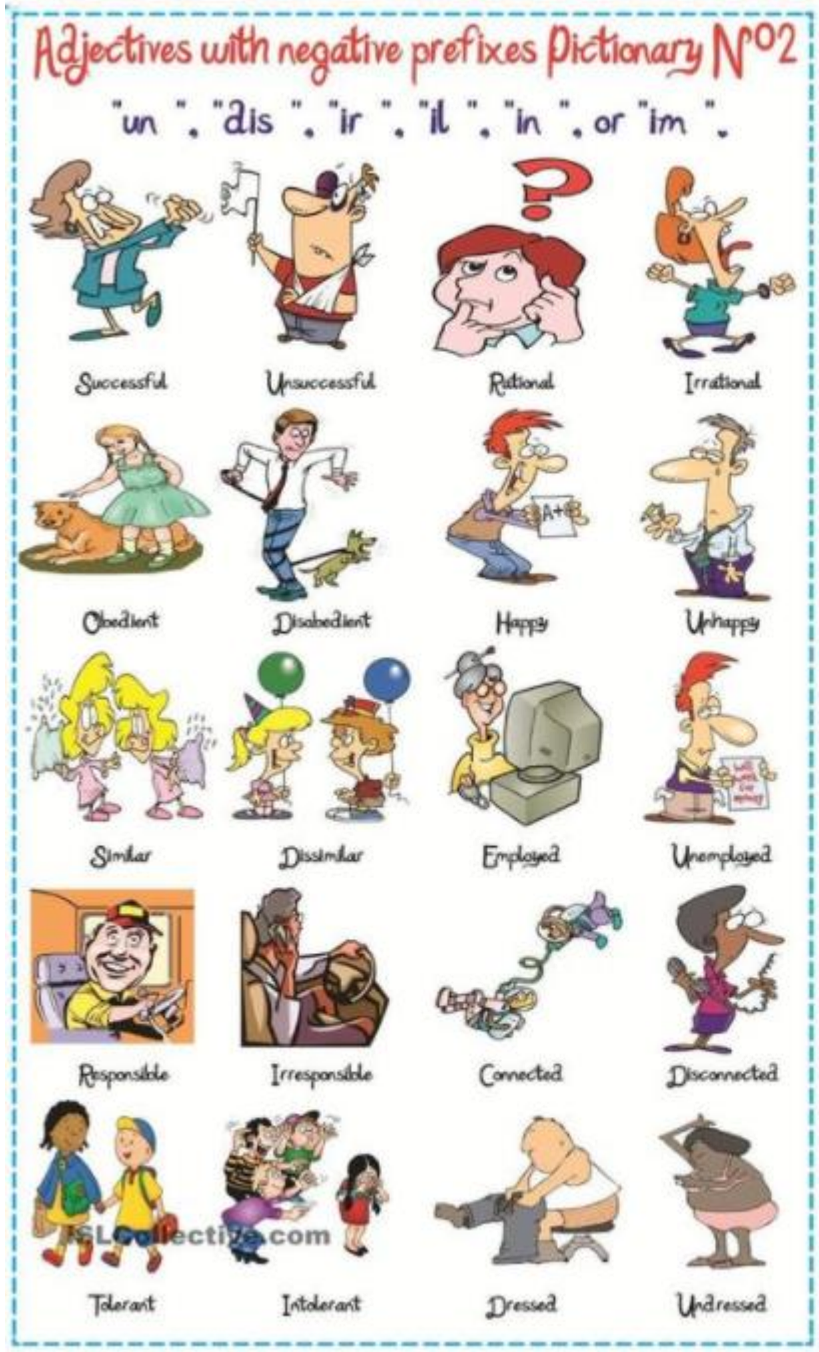
- High correlation between the presence of an adjective and the sentence subjectivity
  - certain adjectives are good indicators of sentiment
  - this does not imply though that other parts of speech do not contribute to expressions of opinion or sentiment

Syntax

- Incorporating syntactic relations within the feature set
  - found to be relevant for short pieces of text (Kudo & Matsumoto)
  - up for debate if this information is helpful

but

  - can provide a basis for including **negation**, collocation and syntactic patterns



Negation

- Handling negation can be important
  - I don’t like this book == I dislike this book != I like this book
- One approach is to initially ignore the negation and the flip the label in the end
  - However, not all appearances of explicit negation terms reverse the polarity of a text segment
    - *No wonder this is considered one of the best* == good regardless of the *No*
  - Negation can be expressed in subtle ways
    - *it avoids all clichés and predictability found in Hollywood movies*

Topic oriented features

- Topics can play a role in the sentiment

  - Walmart reports that profits rose
  - Target reports that profits rose

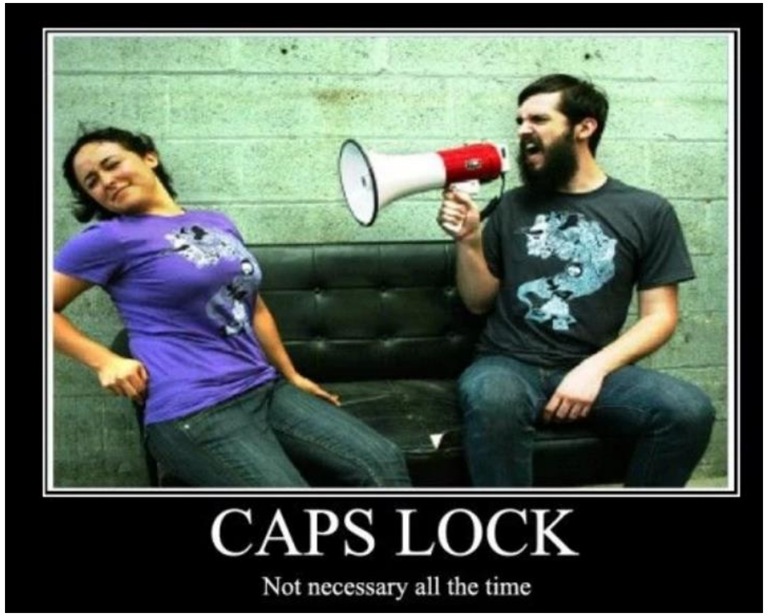
- Indicate completely different types of news (good vs bad) depending on the subject of the document in this case *Walmart*

- Basically we are asking if the sentence is referring to the topic being discussed … sometimes it does but not necessarily

Microblogging features

- Smileys
  - common approach for working with tweets or short texts
    - sentiment mostly succintly represented with the emoticons if they are present

- All caps character representation
  - usually someone is a bit upset if they are using all caps somewhere in the text



Hashtags

- Pre-identifying
  - positive, negative and neutral
  - using them as flags

Positive	#iloveitwhen, #thingsilike, #bestfeeling, #bestfeelingever, #omgthatssotrue, #imthankfulfor, #thingsilove, #success
Negative	#fail, #epicfail, #nevertrust, #worst, #worse, #worstlies, #imtiredof, #itsnotokay, #worstfeeling, #notcute, #somethingaintright, #somethingsnotright, #ihate
Neutral	#job, #tweetajob, #omgfacts, #news, #listeningto, #lastfm, #hiring, #cnn

Table 3: Top positive, negative and neutral hashtags used to create the HASH data set

- treat them as a single feature

- split them into recognizable words and include the individual words in the feature set

Polarity key words

- There seems to be some relation between positive words and positive reviews
  - Number of datasets that identify positive and negative words
  - One question though was :
    - Can humans actually come up with a set of keywords by hand to identify polarity?

Pang et al (2002)

	Features	# of features	frequency or presence?	NB	ME	SVM
(1)	unigrams	16165	freq.	**78.7**	N/A	72.8
(2)	unigrams	"	pres.	81.0	80.4	**82.9**
(3)	unigrams+bigrams	32330	pres.	80.6	80.8	**82.7**
(4)	bigrams	16165	pres.	77.3	**77.4**	77.1
(5)	unigrams+POS	16695	pres.	81.5	80.4	**81.9**
(6)	adjectives	2633	pres.	77.0	**77.7**	75.1
(7)	top 2633 unigrams	2633	pres.	80.3	81.0	**81.4**
(8)	unigrams+position	22430	pres.	81.0	80.1	**81.6**

Two human subjects were asked to pick keywords that would be good indicators of sentiment polarity

	Proposed word lists	Accuracy	Ties
Human 1	positive: *dazzling, brilliant, phenomenal, excellent, fantastic* negative: *suck, terrible, awful, unwatchable, hideous*	58%	75%
Human 2	positive: *gripping, mesmerizing, riveting, spectacular, cool, awesome, thrilling, badass, excellent, moving, exciting* negative: *bad, cliched, sucks, boring, stupid, slow*	64%	39%

Figure 1: Baseline results for human word lists. Data: 700 positive and 700 negative reviews.

Keywords: EmoLex

Both Word-Emotion and Word-Sentiment Association Lexicon

1. NRC Word-Emotion Association Lexicon (also called EmoLex). README. Explore the interactive visualization.
Homepage of the Lexicon.
Also available in over 20 other languages here.

	0.92 (2010)	14,182 unigrams (words)	**sentiments:** negative, positive **emotions:** anger, anticipation, disgust, fear, joy, sadness, surprise, trust	0 (not associated) or 1 (associated)	**Manual:** By crowdsourcing on Mechanical Turk. **Domain:** General
		~25,000 senses*		not associated, weakly, moderately, or strongly associated	

Papers:

Crowdsourcing a Word-Emotion Association Lexicon, Saif Mohammad and Peter Turney, *Computational Intelligence*, 29 (3), 436-465, 2013. Paper (pdf) BibTeX

Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon, Saif Mohammad and Peter Turney, In *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, June 2010, LA, California. Paper (pdf) BibTeX Presentation

http://saifmohammad.com/WebPages/lexicons.html

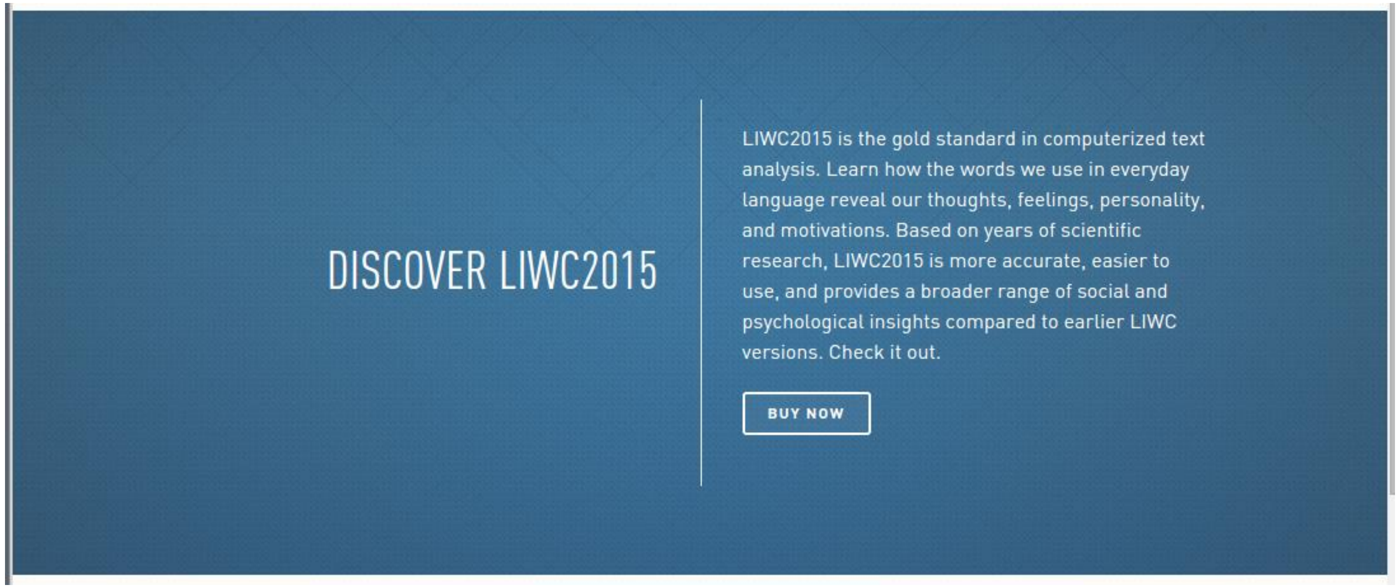
Keywords: Hashtag sentiment lexicon

3. NRC Twitter Sentiment Lexicons (NRC Hashtag Sentiment Lexicons and Sentiment140 Lexicons)					
a. NRC Hashtag Sentiment Lexicon					
	1.0 (2013)	54,129 unigrams 316,531 bigrams 308,808 pairs	**sentiments:** negative, positive	Real-valued score between -∞ (most negative) to ∞ (most positive)	**Automatic**: From tweets with sentiment word hashtags. **Domain:** Twitter
b. NRC Hashtag Affirmative Context Sentiment Lexicon and NRC Hashtag Negated Context Sentiment Lexicon					
	1.0 (2014)	Affirmative contexts: 36,357 unigrams Negated contexts: 7,592 unigrams Affirmative contexts: 159,479 bigrams Negated contexts: 23,875 bigrams	**sentiments:** negative, positive	Real-valued score between -∞ (most negative) to ∞ (most positive)	**Automatic:** From tweets with sentiment word hashtags. Separate entries for affirmative and negated contexts. **Domain:** Twitter

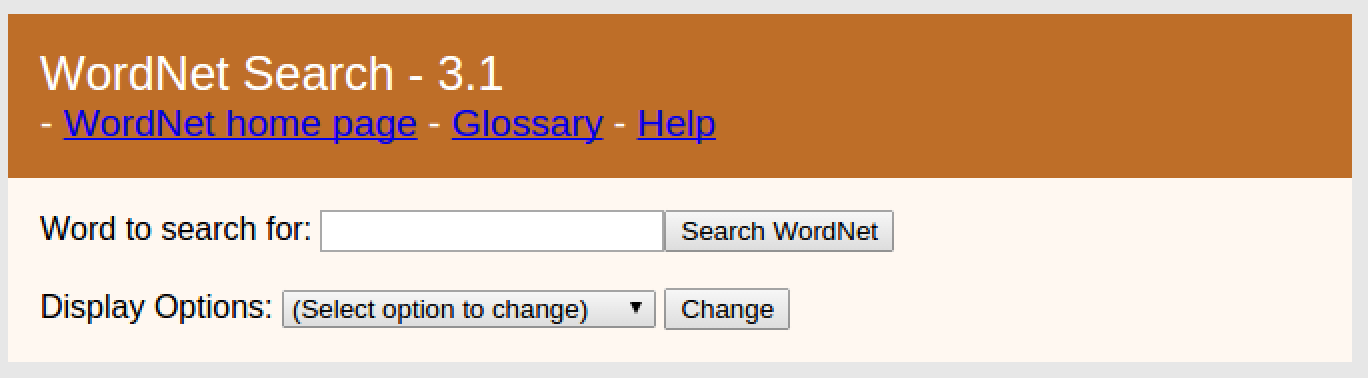
http://saifmohammad.com/WebPages/lexicons.html

Keywords: LIWC

- Linguistic Inquiry and Word Count
  - Similar to General Inquirer
- Counts words belonging to categories including positive and negative
- http://liwc.wpengine.com/



Keywords: WordNet



- We know what this is:
  - Nouns, verbs and adjectives are grouped into synsets
- Not specifically sentiment oriented but has been used to derive sentiment related information (Hu & Liu, 2004)
- https://wordnet.princeton.edu/

Keywords: SentiWordNet

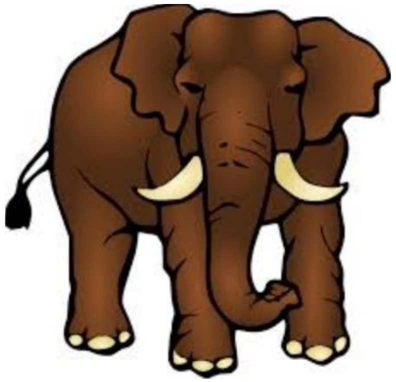
- A lexical resource for opinion mining
  - assigns three probability like scores to each WordNet synset for
    - objectivity, positivity and negativity
  - scores are based on an estimation algorithm (not manual)
  - powerful resource for estimating the sentiment of individual words
  - needs linguistic processing of source text to match words to synsets
- http://sentiwordnet.isti.cnr.it/

Feature-based

What difficulty do we have with these feature-based representations?

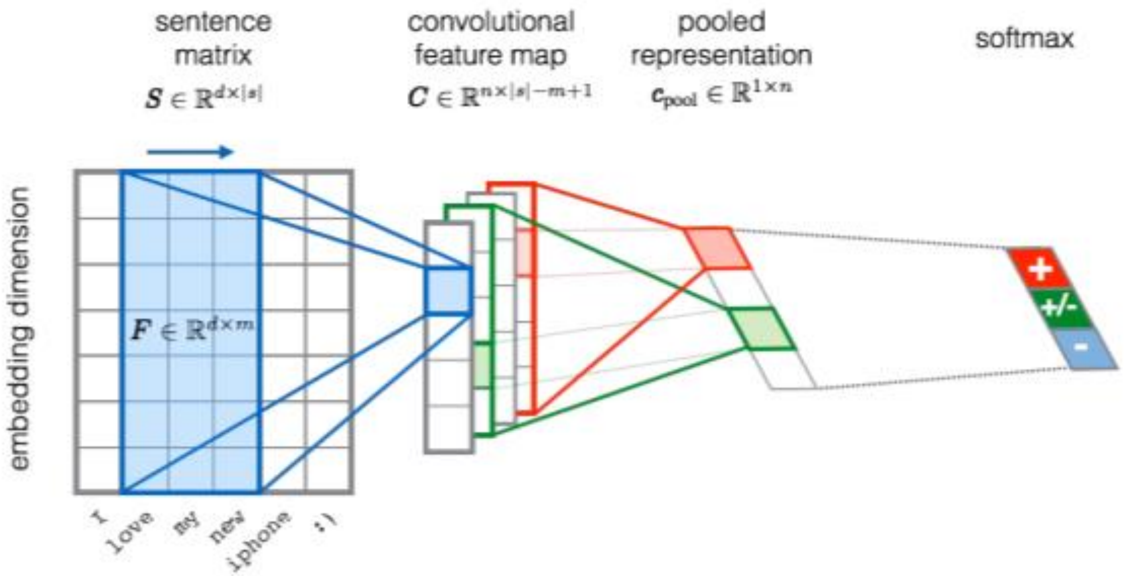
Feature-based

What difficulty do we have with these feature-based representations?

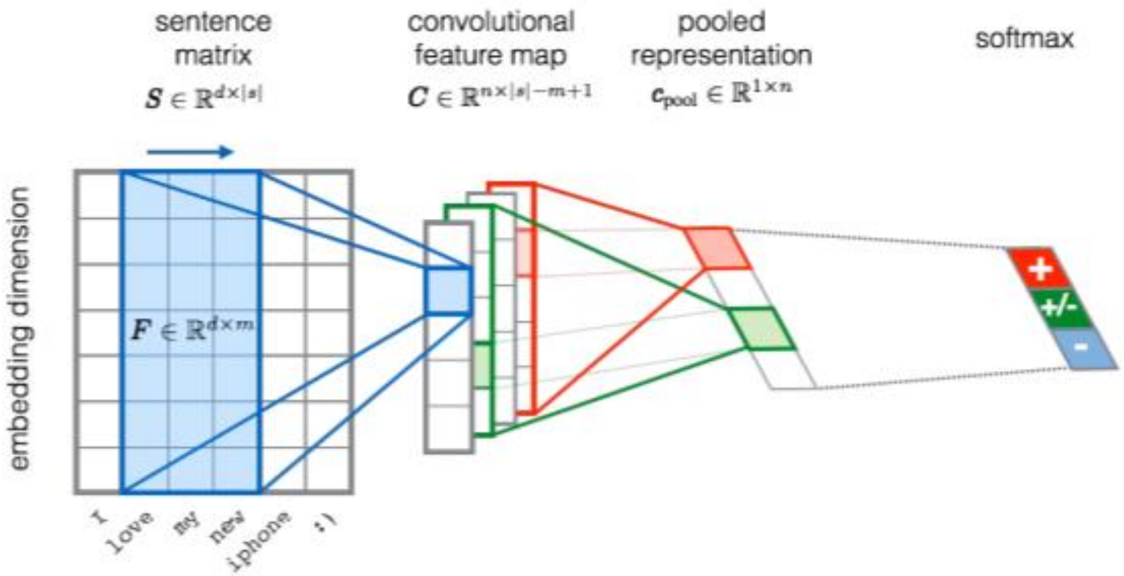


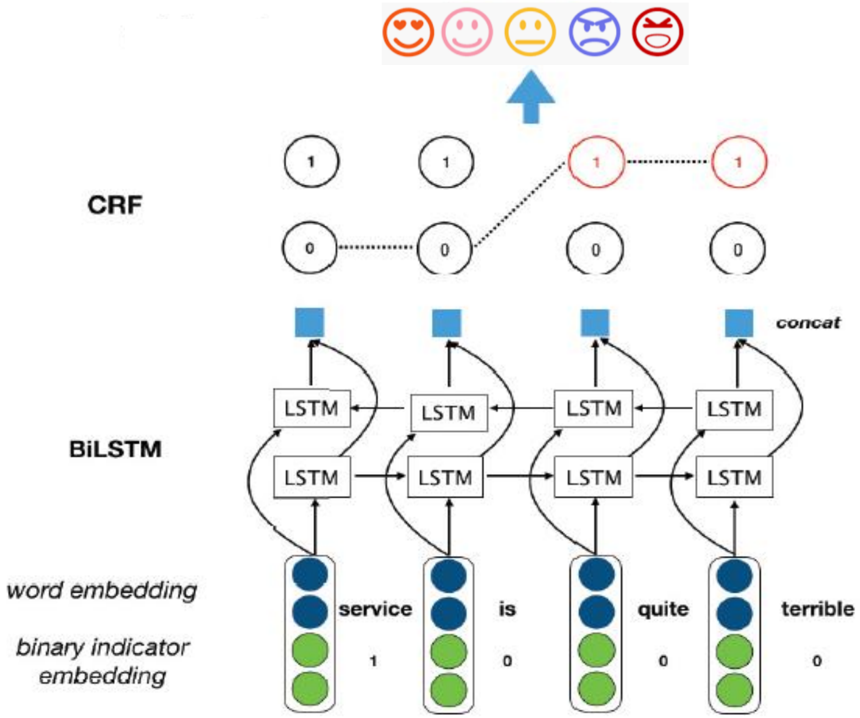
Featureless

- Deep Learning Algorithm:
    - cNN

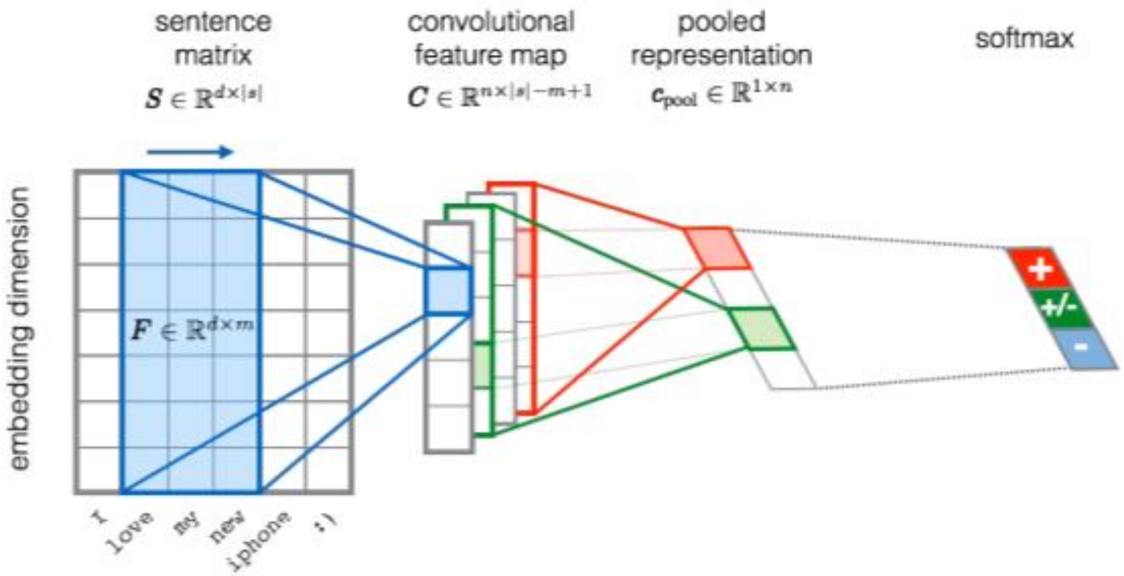


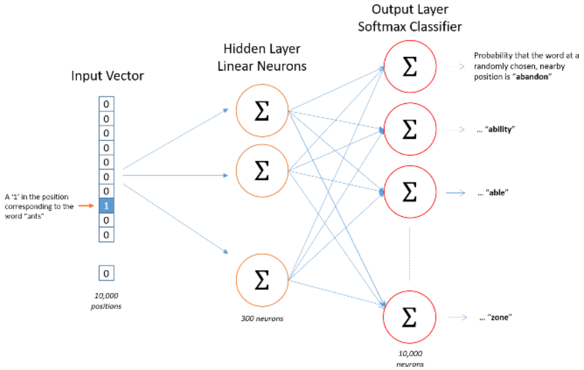
- Deep Learning Algorithm:
  - cNN
  - bi-LSTM+CRF

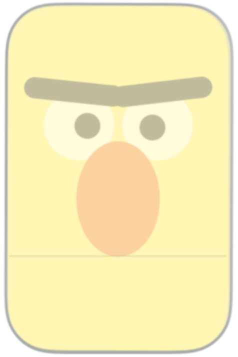


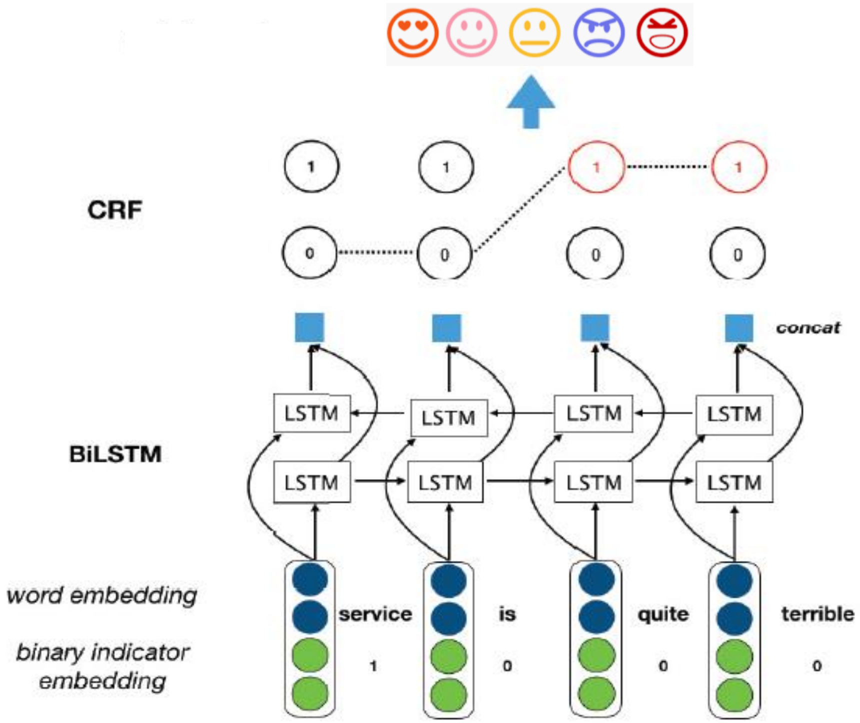


- Deep Learning Algorithm:
    - cNN
    - bi-LSTM
- Features:
    - word embeddings
        - Glove
        - W2V
        - BERT
        - ELMO
        - ERNIE

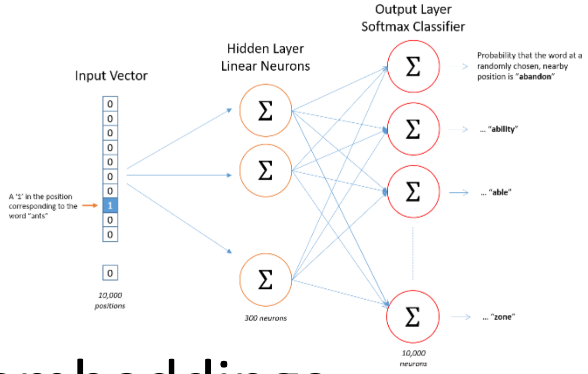


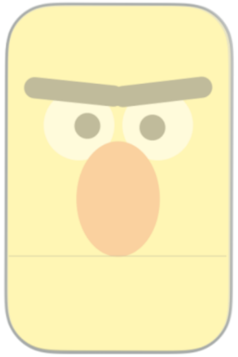


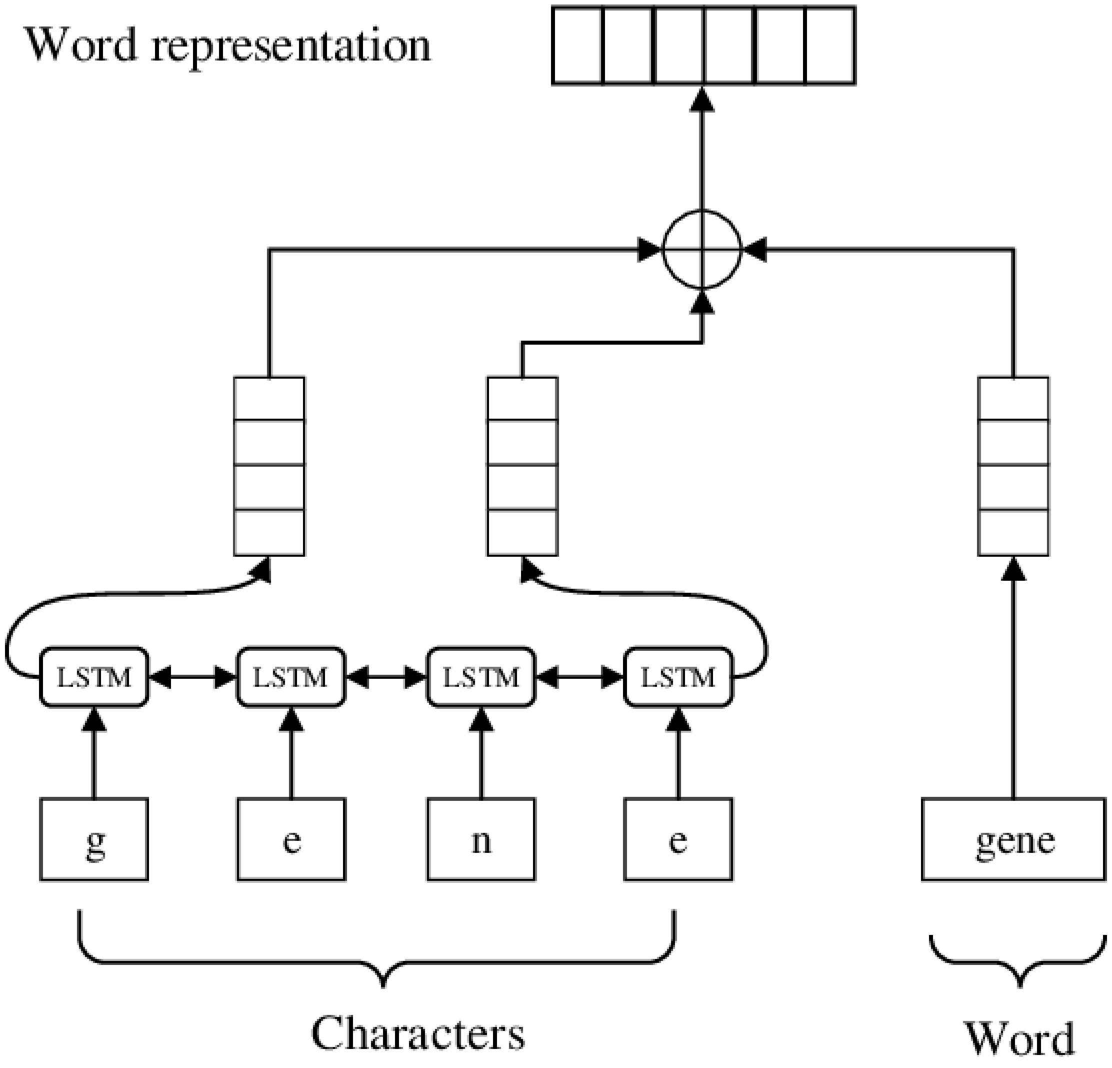




- Deep Learning Algorithm:
    - cNN
    - bi-LSTM
  - Features:
    - word embeddings
      - Glove
      - W2V
      - BERT
      - ELMO
      - ERNIE
    - character embeddings







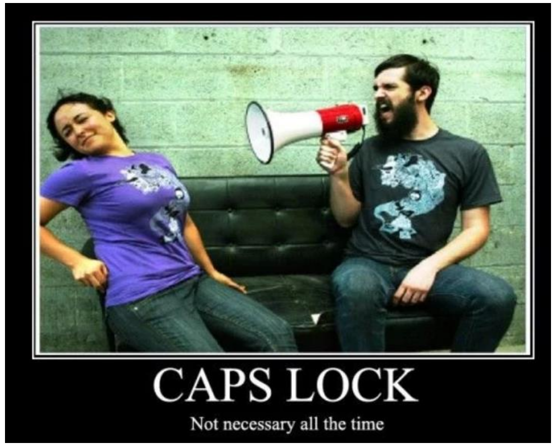
Featureless

What is the problem though with these featureless representations for sentiment analysis?

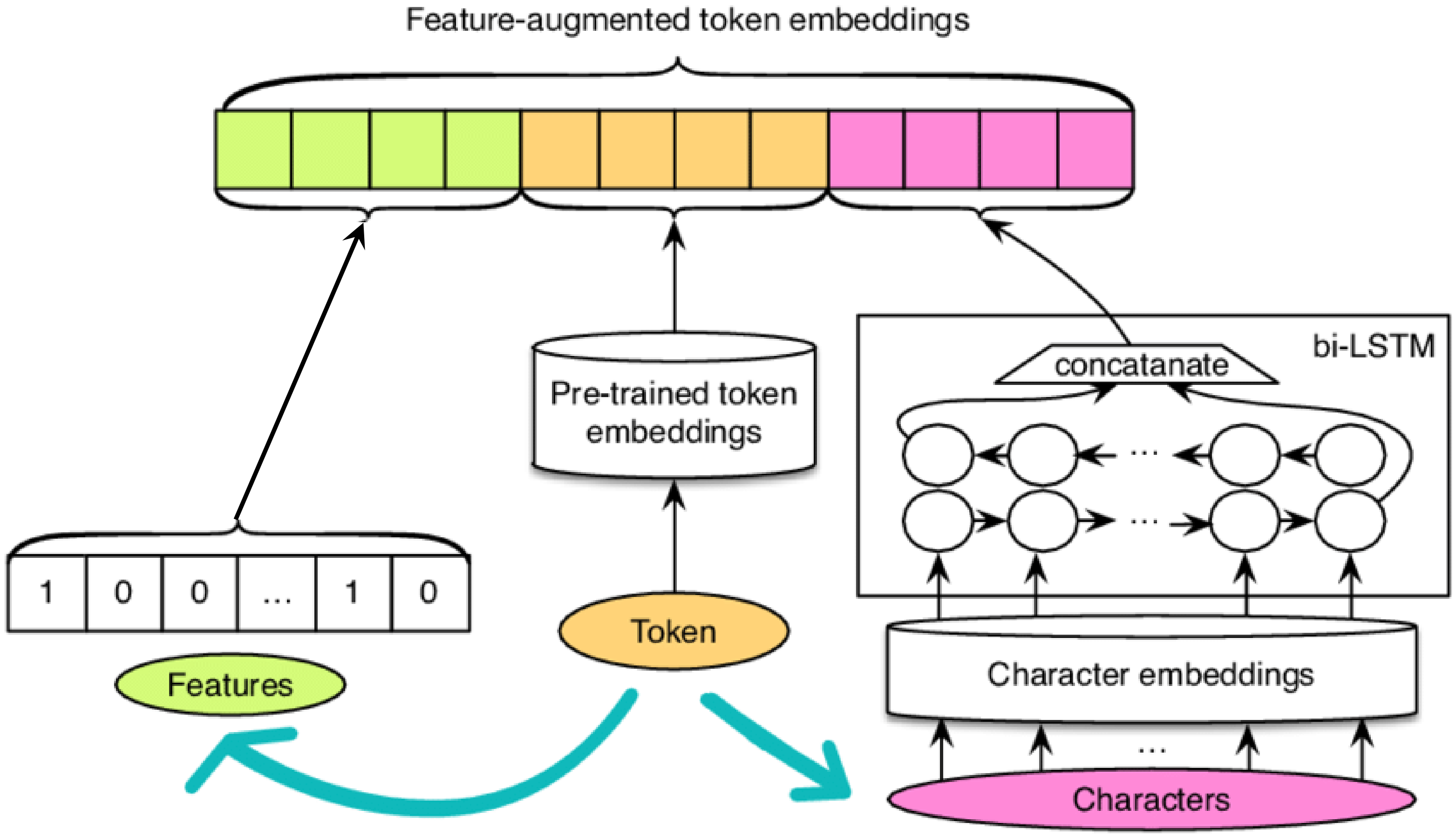
Featureless

What is the problem though with these featureless representations for sentiment analysis?

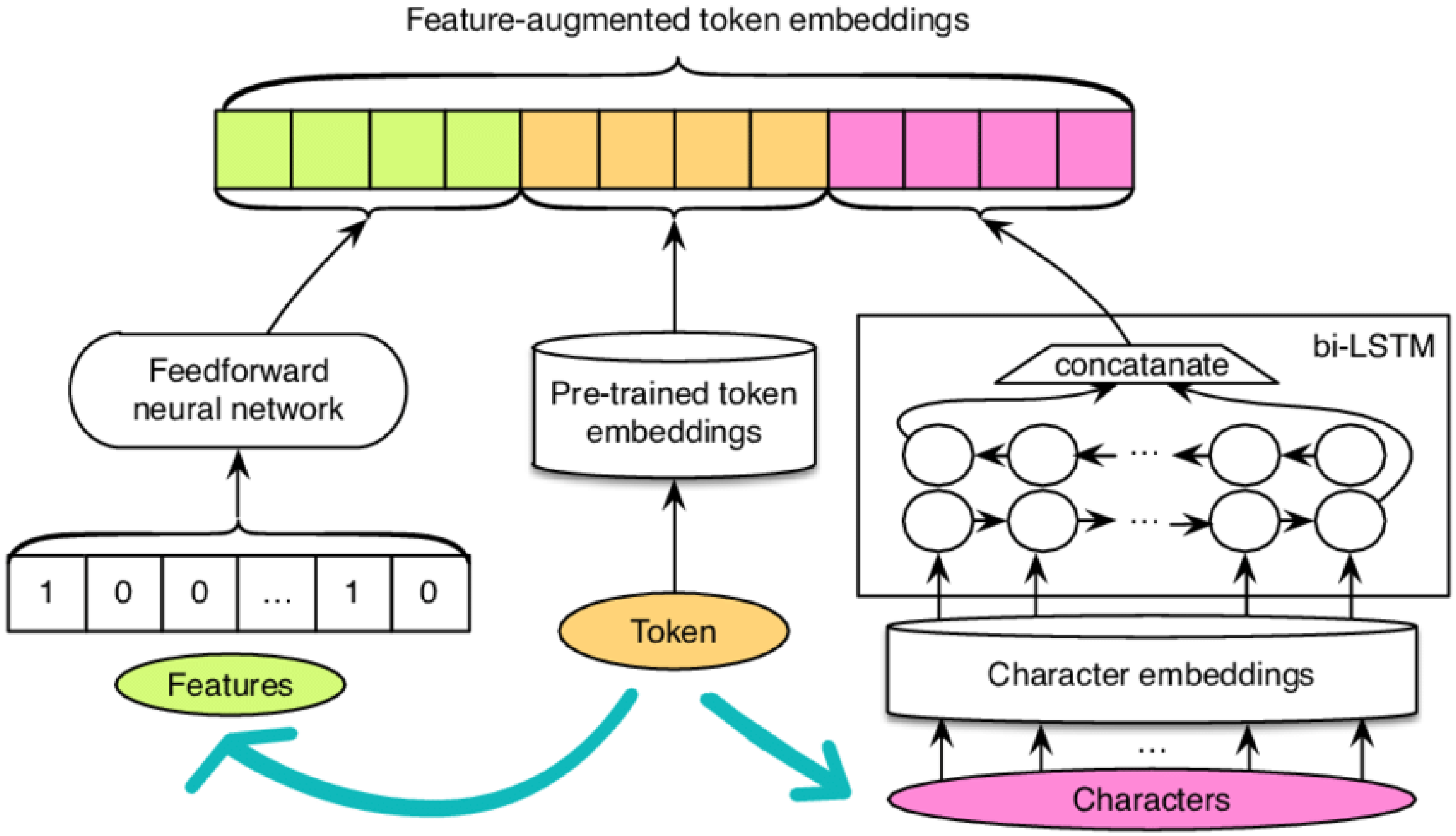
Losing structural knowledge



Combining feature-based and featureless representations



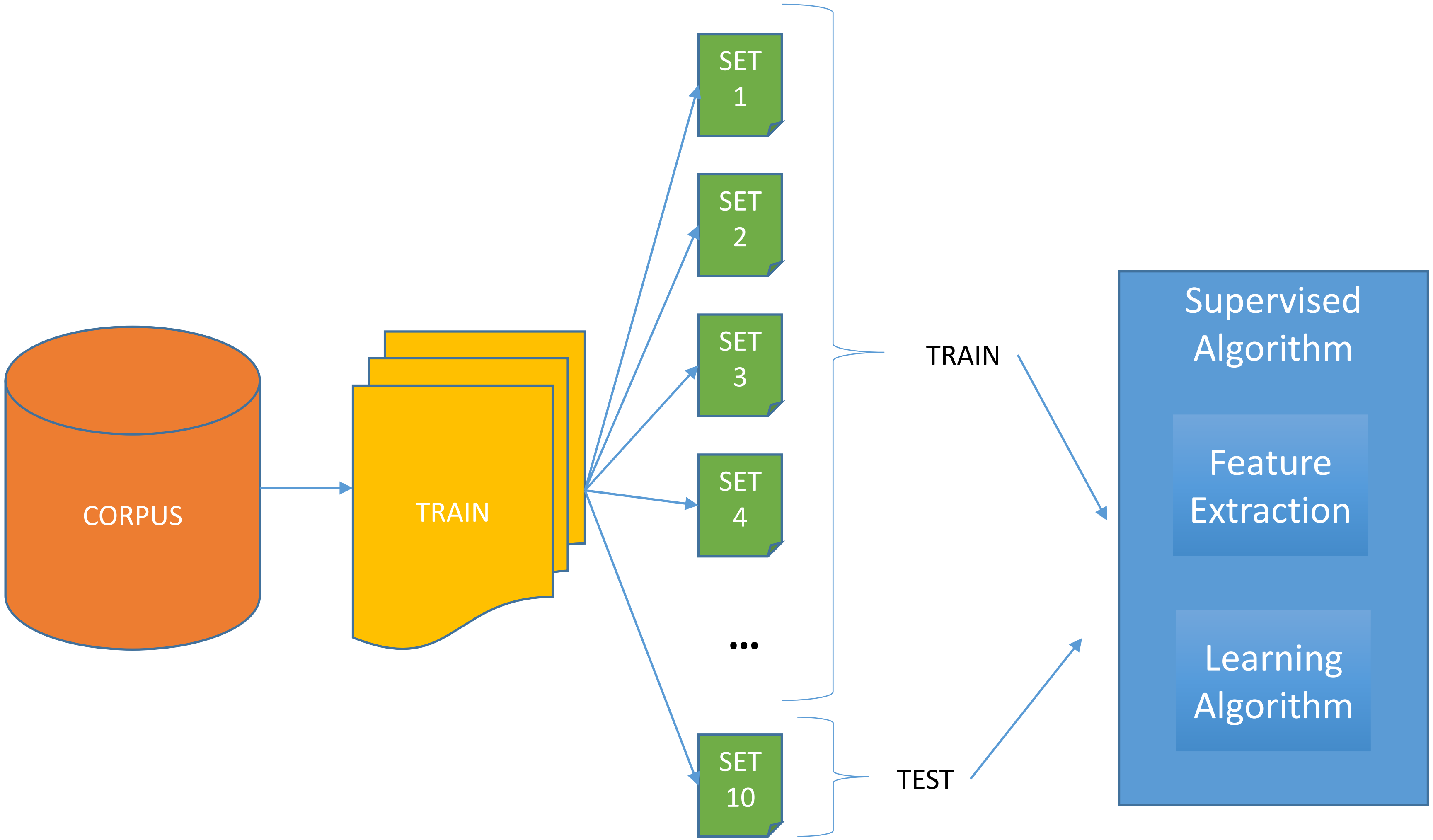
Combining feature-based and featureless representations



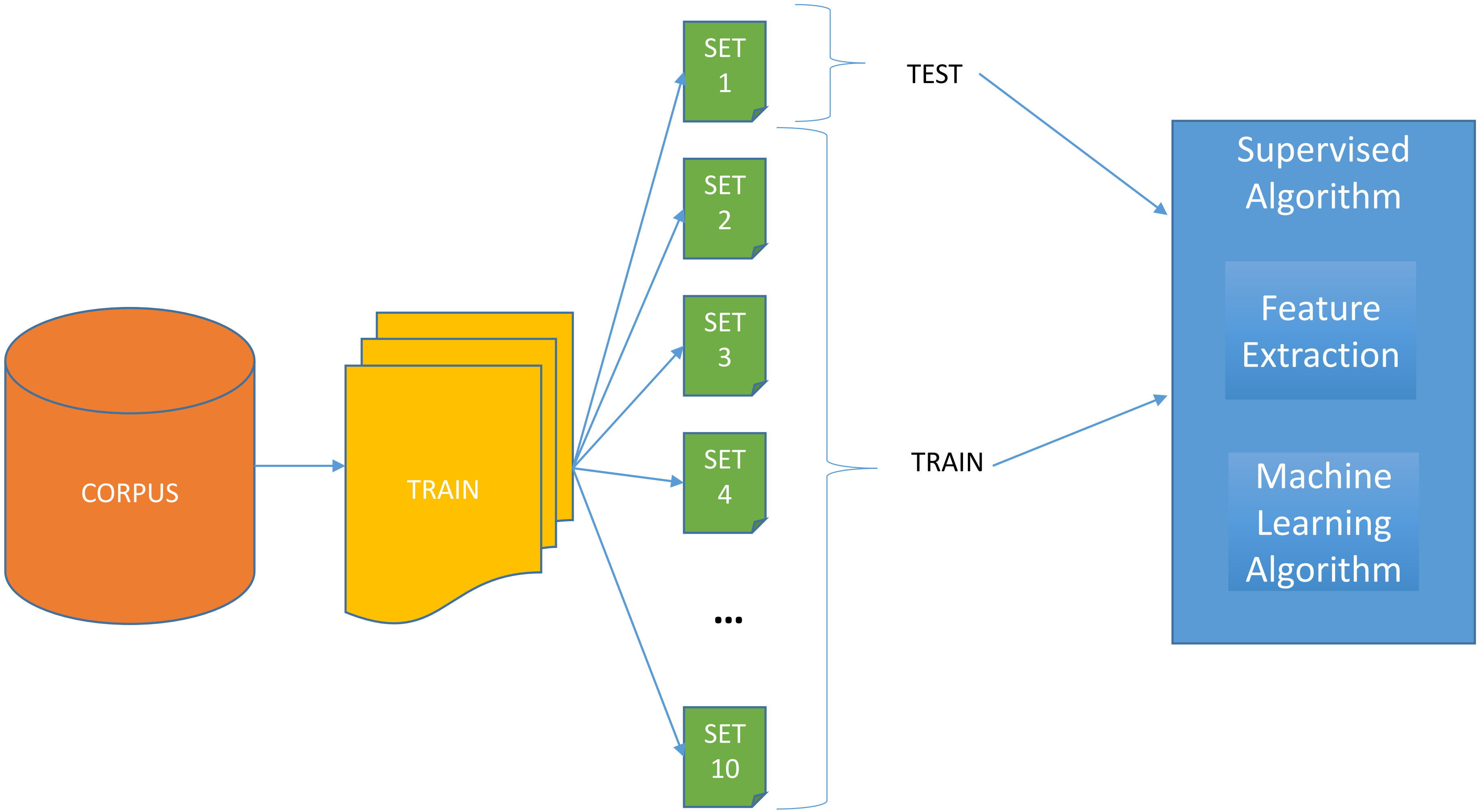
Questions, concerns, queries?

Evaluation methodologies

10 Fold Cross Validation Evaluation Methodology

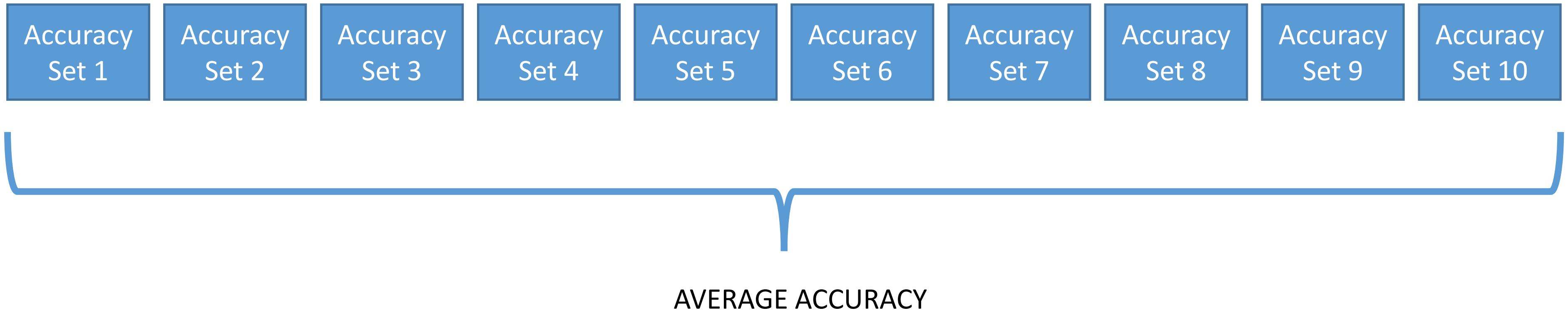


10 Fold Cross Validation Evaluation Methodology



Cycle Through

- Each file is used as a test once and only once
  - With the remaining used as train



Evaluation Metrics

- Accuracy
- Precision
- Recall
- F-measure

Accuracy

statistical measure of how well the classification went

- $accuracy = \frac{number\ correctly\ tagged\ words}{total\ number\ of\ words}$

precision, recall and f-measure

- Precision
  - How accurate the system was at labeling the instances it labeled
- Recall
  - How well the system identified all the instances it was suppose to
- Fmeasure
  - The harmonic mean between the two

- $precision = \frac{tp}{tp+fp}$
- $recall = \frac{tp}{tp+fn}$
- $fmeasure = 2 * \frac{precision * recall}{precision + recall}$

- tp = true positive
- fp = false positive
- fn = false negative
- fp = false positive

Datasets for semantic analysis learning

Pang & Lee datasets

- Movie review polarity datasets
- Sentiment scale datasets
- Subjectivity datasets

Movie Review Data

This page is a distribution site for movie-review data for use in sentiment-analysis experiments. Available are collections of movie-review documents labeled with respect to their overall *sentiment polarity* (positive or negative) or *subjective rating* (e.g., "two and a half stars") and sentences labeled with respect to their *subjectivity status* (subjective or objective) or *polarity*. These data sets were introduced in the following papers:

- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, Thumbs up? Sentiment Classification using Machine Learning Techniques, *Proceedings of EMNLP 2002*.
- Bo Pang and Lillian Lee, A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, *Proceedings of ACL 2004*.
- Bo Pang and Lillian Lee, Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, *Proceedings of ACL 2005*.

If you have results to report on these corpora, please send email to Bo Pang and/or Lillian Lee so we can add you to our **list of other papers using this data.** Thanks! Rationale: we're (admittedly haphazardly and only occcasionally) maintaining that list for the purposes of facilitating comparison of results.

Sentiment polarity datasets

- polarity dataset v2.0 (3.0Mb) (includes README v2.0): 1000 positive and 1000 negative processed reviews. Introduced in Pang/Lee ACL 2004. Released June 2004.
- Pool of 27886 unprocessed html files (81.1Mb) from which the polarity dataset v2.0 was derived. (This file is identical to movie.zip from data release v1.0.)
- sentence polarity dataset v1.0 (includes sentence polarity dataset README v1.0: 5331 positive and 5331 negative processed sentences / snippets. Introduced in Pang/Lee ACL 2005. Released July 2005.

- archive:
  - polarity dataset v1.0 (2.8Mb) (includes README): 700 positive and 700 negative processed reviews. Released July 2002.
  - polarity dataset v1.1 (2.2Mb) (includes README.1.1): approximately 700 positive and 700 negative processed reviews. Released November 2002. This alternative version was created by Nathan Treloar, who removed a few non-English/incomplete reviews and changing some of the labels (judging some polarities to be different from the original author's rating). The complete list of changes made to v1.1 can be found in diff.txt.
  - polarity dataset v0.9 (2.8Mb) (includes a README):. 700 positive and 700 negative processed reviews. Introduced in Pang/Lee/Vaithyanathan EMNLP 2002. Released July 2002. Please read the "Rating Information - WARNING" section of the README.
  - movie.zip (81.1Mb): all html files we collected from the IMDb archive.

Subjectivity datasets

- subjectivity dataset v1.0 (508K) (includes subjectivity README v1.0): 5000 subjective and 5000 objective processed sentences. Introduced in Pang/Lee ACL 2004. Released June 2004.
- Pool of unprocessed source documents (9.3Mb) from which the sentences in the subjectivity dataset v1.0 were extracted. **Note**: On April 2, 2012, we replaced the original gzipped tarball with one in which the subjective files are now in the correct directory (so that the subjectivity directory is no longer empty; the subjective files were mistakenly placed in the wrong directory, although distinguishable by their different naming scheme).

- http://www.cs.cornell.edu/people/pabo/movie-review-data/

Blitzer et al Multi-domain sentiment dataset

- Reviews from amazon.com from any product types
- Includes the star ratings
- Divided into positive and negative ratings

Multi-Domain Sentiment Dataset (version 2.0)

This sentiment dataset supersedes the previous data (still available here).

Link to download the data:
[unprocessed.tar.gz] (1.5 G)
[processed_acl.tar.gz] (19 M)
[processed_stars.tar.gz] (33 M)

This sentiment dataset has been used in several papers:

John Blitzer, Mark Dredze, Fernando Pereira. Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. Association of Computational Linguistics (ACL), 2007. [PDF]

John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jenn Wortman. Learning Bounds for Domain Adaptation. Neural Information Processing Systems (NIPS), 2008. [PDF]

Mark Dredze, Koby Crammer, and Fernando Pereira. Confidence-Weighted Linear Classification. International Conference on Machine Learning (ICML), 2008. [PDF]

Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Domain Adaptation with Multiple Sources. Neural Information Processing Systems (NIPS), 2009.

- http://www.cs.jhu.edu/~mdredze/datasets/sentiment/

MPQA Opinion Corpus

- Multi-perspective Question Answering (Stoyanov et al 2005)
- News articles and other text documents are manually annotated for opinions and other private states
  - ie – beliefs, emotions, sentiments, speculations
- 692 documents (15,802 sentences)
- http://mpqa.cs.pitt.edu/

- MPQA Opinion Corpus

The MPQA Opinion Corpus contains news articles from a wide variety of news sources manually annotated for opinions and other private states (i.e., beliefs, emotions, sentiments, speculations, etc.). To download the MPQA Opinion Corpus click here.

For **sample documents and instructions for MPQA annotation in GATE**, click here. *Updated July 2011.*

To learn more about the subjectivity and sentiment research that produced MPQA, please refer to the following publications:

Janyce Wiebe, Theresa Wilson , and Claire Cardie (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation, volume 39, issue 2-3, pp. 165-210.*

Theresa Wilson (2008). Fine-Grained Subjectivity Analysis. *PhD Dissertation, Intelligent Systems Program, University of Pittsburgh.*

Additional Training/explanatory materials coming soon.

Thomas, Pang and Lee, 2006

- Congressional speech data
  - transcripts of floor debates on policy

- http://www.cs.cornell.edu/home/llee/data/convote.html

Congressional speech data

This page is a distribution site for a congressional-speech corpus and related extracted information. This data includes speeches as individual documents, together with:

- automatically-derived labels for whether the speaker supported or opposed the legislation discussed in the debate the speech appears in, allowing for experiments with this kind of sentiment analysis
  - We also maintain and distribute another corpus of data suitable for work in sentiment analysis, the Cornell movie-review data set.
- indications of which "debate" each speech comes from, allowing for consideration of conversational structure
- indications of by-name references between speakers, and the scores that our agreement/disagreement classifier(s) automatically assigned to such references, allowing for experiments on agreement classification if one assigns "true" labels from the support/oppose labels assigned to the pair of speakers in question
- the edge weights and other information we derived to create the graphs we used for our experiments upon this data, facilitating implementation of alternative graph-based methods upon the graphs we constructed

If you have used this data, we would appreciate hearing about it (Lillian Lee is our designated contact person); a list of those papers we know about can be found below.

Sentiment analysis on twitter

SemEval-2017 Task 4: Sentiment Analysis in Twitter

Sara Rosenthal, Noura Farra, Preslav Nakov

Abstract

This paper describes the fifth year of the Sentiment Analysis in Twitter task. SemEval-2017 Task 4 continues with a rerun of the subtasks of SemEval-2016 Task 4, which include identifying the overall sentiment of the tweet, sentiment towards a topic with classification on a two-point and on a five-point ordinal scale, and quantification of the distribution of sentiment towards a topic across a number of tweets: again on a two-point and on a five-point ordinal scale. Compared to 2016, we made two changes: (i) we introduced a new language, Arabic, for all subtasks, and (ii) we made available information from the profiles of the Twitter users who posted the target tweets. The task continues to be very popular, with a total of 48 teams participating this year.

Creating sentiment-oriented datasets

- Self-annotated data
  - data has built in ordinal or binary labeling of some kind to complement the natural language text – ideally by the author of the text
  - Examples
    - amazon reviews
    - pitchfork.com record reviews

- Hand-annotated data
  - annotated independently
  - labour intensive

Inter-annotator agreement

- Hand annotated sentiment data can very in reliability
- Inter-annotator agreement is the degree to which multiple human annotators arrive at the same annotations when confronted with the same natural language text
- Represents a theoretical upper bound for the classification
- Cohen’s kappa
  - measure of agreement between two raters
    - pairwise average for multiple

To sum up

- Sentiment analysis is a difficult task
- The difficulty increases with the nuance and complexity of opinions being expressed

PA 5

Notes:

- Go to the programming assignment on blackboard
- Show the data