Lecture 7:

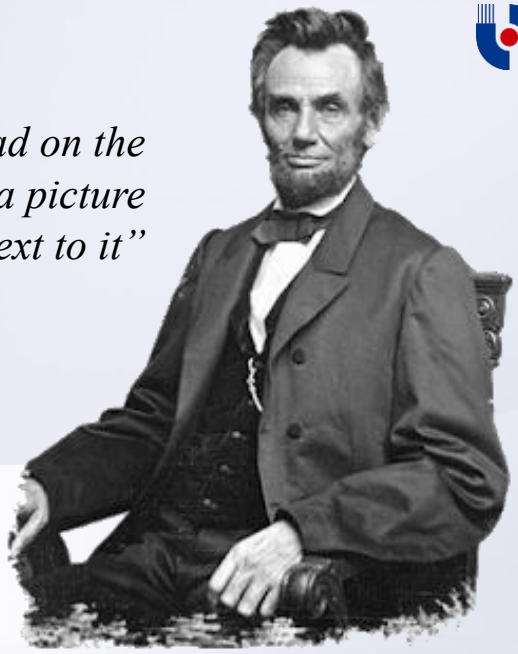
Document Sentiment Classification

Social Media Computing
CDS - 6344
2024 - 2025

ATM University

"Don't believe everything you read on the Internet just because there's a picture of someone famous with a quote next to it"

- Ibrahim Limkan



Sentiment Analysis Tasks

- Sentiment Classification : classify a piece of text based on whether it expresses a Positive / Neutral / Negative sentiment
- Sentiment Lexicon Generation : determine whether a word / multiword conveys a Positive, Neutral, or Negative sentiment
- Sentiment Quantification : given a set of texts, estimate the prevalence of different Positive, Neutral, Negative sentiments
- Opinion Extraction (Fine-Grained SA) : given an opinion-laden sentence, identify the holder of the opinion, its object, its polarity, the strength of this polarity, the type of opinion
- Aspect-Based Sentiment Extraction : given an opinion-laden text about an object, estimate the sentiments conveyed by the text concerning different aspects of the object

Levels of Analysis and Outcomes

- It is easier to reference the sentiment analysis tasks based on the scope it is applied to:
  - document level,
  - sentence level ← fine-grained
  - aspect level
- Both the document level and sentence level classifications are already highly challenging. The aspect level is even more difficult

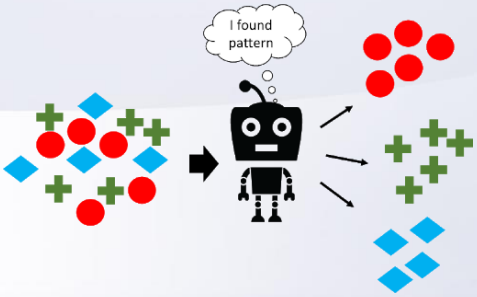
- Let's focus on document level for today's content

Document Level Sentiment Analysis

- Document-level sentiment classification : Process to classify whether a whole opinion document express either a positive/negative sentiment
  - For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product
- It is basically a text classification problem → This means any conventional supervised learning algorithm can be used
- Implicitly assumes that each document expresses opinions on a single entity (e.g., a single product or service).
- Thus it is not applicable to documents that evaluate or compare multiple entities, for which more fine-grained analysis is needed

Document Sentiment Categories

- There are in fact two popular formulations of document-level sentiment analysis based on the type of values that [s] takes in the quintuple
  - If [s] takes on categorical values (1-5 scale, yes/no) then it becomes a classification problem
  - If [s] takes numerical values (ranges) then it becomes a regression problem



Goals / Assumptions

- For document level SA, the obvious assumption is the doc is written by a single person and expresses opinion/sentiment on a single entity
  - i.e. Goal: (_,general,S,_,_) where e,a,h and t are often ignored
  - Product/Service reviews usually satisfy the assumptions but many postings/blogs will not
  - Many document SA operations rely on supervised learning to detect sentiment
    → i.e. naïve bayes, max entropy, SVM etc.

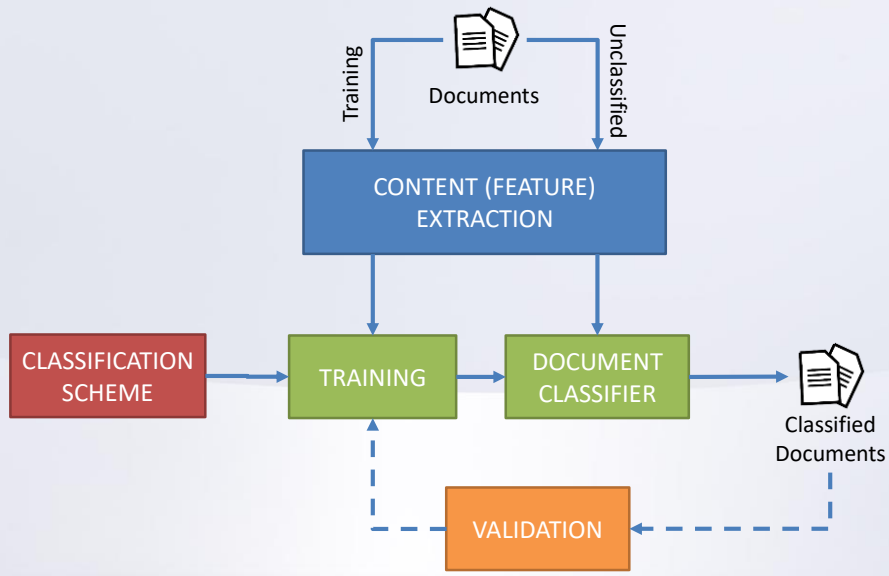
Supervised Sentiment Classification

- Majority sentiment classification is usually formulated as a two-class classification problem: *positive* and *negative*
  - The training and testing data used can be documents like product reviews
  - A review with 4 or 5 stars is considered a *positive* review, and a review with 1 to 2 stars is considered a *negative* review. Most research papers do not use the neutral class (3-star ratings) to make the classification problem easier
- Conventional classification of documents look for topics whereas for sentiment analysis, opinion words take importance
- Two popular approaches (1) standard supervised ML algorithm and (2) using sentiment classification method

DOCUMENT SENTIMENT CLASSIFICATION

SUPERVISED LEARNING METHODS

TYPICAL SUPERVISED SA ARCHITECTURE



Training and Test Data

- Training of unsupervised classification algorithms to identify opinions usually require training data and test data to validate
- Depending on domain, the data could naturally have definition for mapping to sentiment, e.g.
  - Movie reviews with star ratings
    - 4-5 stars as positive
    - 1-2 stars as negative
    - 3 stars are neutral → ignored
- Once trained, the classifier then can be applied to unclassified documents of similar nature

Supervised Document Classification

- In order to be input to a classifier, all training/unlabeled documents are converted into vectors in a common vector space → called features (terms) and the number of features is the dimensionality of the vector space
- Typical choice is to make the set of features coincide with the set of words that occur in the training set → usually preceded with stop word removal, stemming, lemmatization to improve classification outcomes

- Like most supervised learning applications, the key for sentiment classification is the engineering of effective features. Examples
  - TF-IDF
  - Part-of-Speech
  - Sentiment words/phrases
  - Rules of opinion
  - Sentiment shifters
  - Syntactic dependency

TF-IDF

- TF-IDF is a statistical measure that evaluates how relevant a word is to a document in a collection of documents. Takes two metrics to derive the measure;
  - how many times a word appears in a document (term frequency) and the inverse document frequency of the word across a set of documents
  - if the word is very common and appears in many documents, this number will approach 0. Otherwise, it will approach 1.
- Example; In a document with 1000 words, the term 'MMU' appears 3 times thus the TF = 3/1000 = 0.003. In a collection of 10million related documents, the term appears 1000 times, thus IDF = log(10000000/1000) = 4.

  Therefore the term 'MMU' has TD-IDF measure of 4 x 0.003 = 0.012

Part-Of-Speech

- POS (grammatical) tagging is the process of marking/tagging words found in text as corresponding to a particular part-of-speech
  - Similar to knowing normal sentence components (i.e. subject, predicate, verb, noun, adverb etc) but at for more detail
- Helps to locate components for sentiments (e.g. adjectives = intensity, subjects = holder etc)
- POS-tagging algorithms fall into two groups : rule-based and stochastic.
- The outcome tags depend on which list is used (there are many)

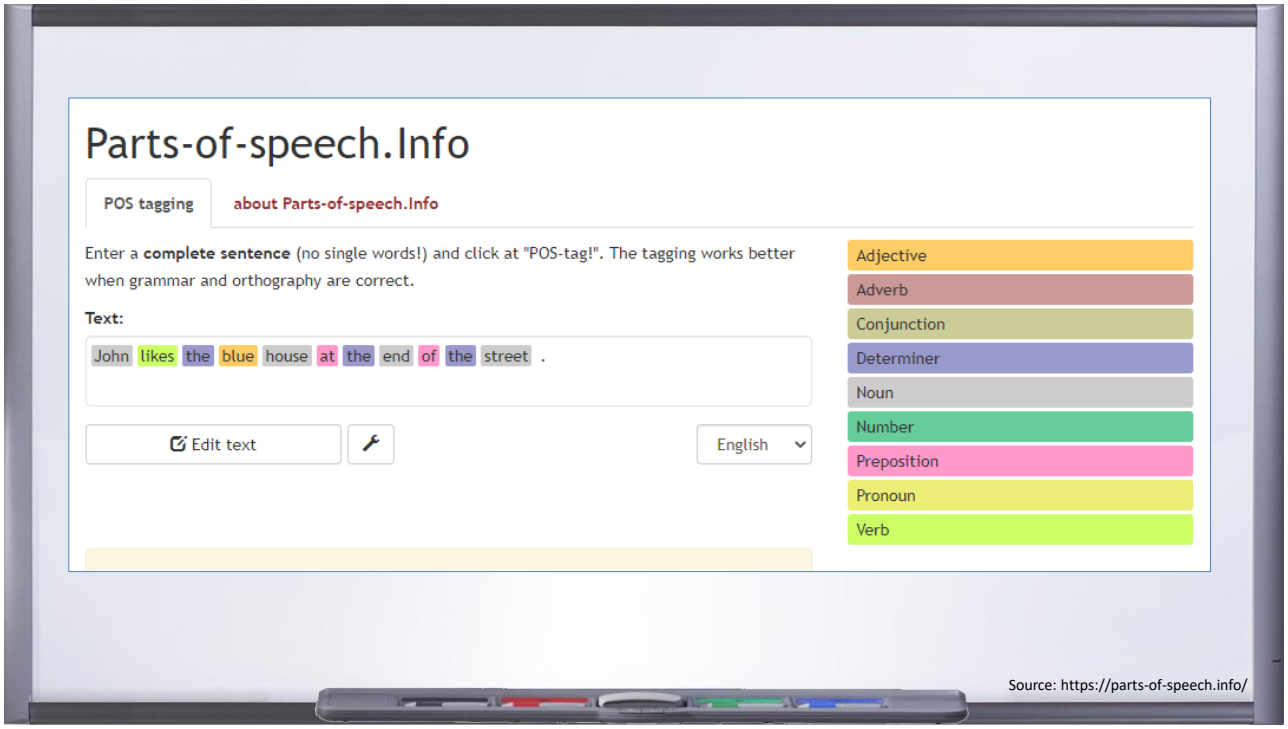
Tag	Description	Tag	Description
CC	Coordinating conjunction	PRP$	Possessive pronoun
CD	Cardinal number	RB	Adverb
DT	Determiner	RBR	Adverb, comparative
EX	Existential *there*	RBS	Adverb, superlative
FW	Foreign word	RP	Particle
IN	Preposition or subordinating conjunction	SYM	Symbol
JJ	Adjective	TO	*to*
JJR	Adjective, comparative	UH	Interjection
JJS	Adjective, superlative	VB	Verb, base form
LS	List item marker	VBD	Verb, past tense
MD	Modal	VBG	Verb, gerund or present participle
NN	Noun, singular or mass	VBN	Verb, past participle
NNS	Noun, plural	VBP	Verb, non-3rd person singular present
NNP	Proper noun, singular	VBZ	Verb, 3rd person singular present
NNPS	Proper noun, plural	WDT	Wh-determiner
PDT	Predeterminer	WP	Wh-pronoun
POS	Possessive ending	WP$	Possessive wh-pronoun
PRP	Personal pronoun	WRB	Wh-adverb

Penn Treebank

Open class words	Closed class words	Other
ADJ	ADP	PUNCT
ADV	AUX	SYM
INTJ	CCONJ	X
NOUN	DET	
PROPN	NUM	
VERB	PART	
	PRON	
	SCONJ	

Universal POS

Source https://repository.upenn.edu/cgi/viewcontent.cgi?article=1603&context=cis_reports
Source : https://universaldependencies.org/u/pos/



Source: https://parts-of-speech.info/

Sentiment Words/Phrases

- Sentiment words are natural features as they are words in a language for expressing positive or negative sentiments → e.g. good, wonderful, amazing, poor, terrible, horrible, disgusting etc.
- Most sentiment words are adjectives or adverbs
  - Sometimes nouns (blue, junk) and verbs (hate, love) can also invoke sentiment
- Multiple words in a phrase/idiom can also represent sentiment
  - "costs an arm and a leg"
  - "happy birthday to you"
  - "walking on sunshine"
  - "heart-breaking"
  - "crash and burn"

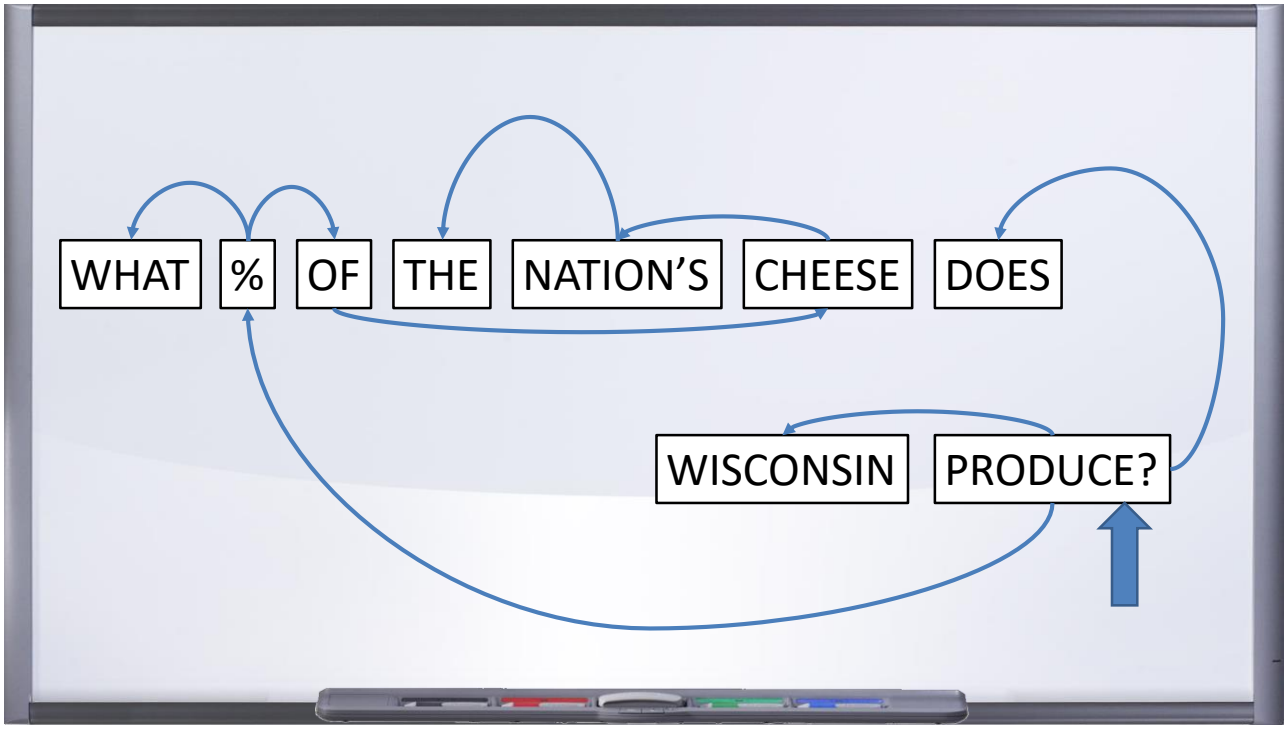
Rules of Opinion / Sentiment Shifter

- Aside from sentiment words and phrases, there are also language constructs (e.g. syntax) that can imply sentiment → these are opinion rules
  - To be covered in further detail in later lecture

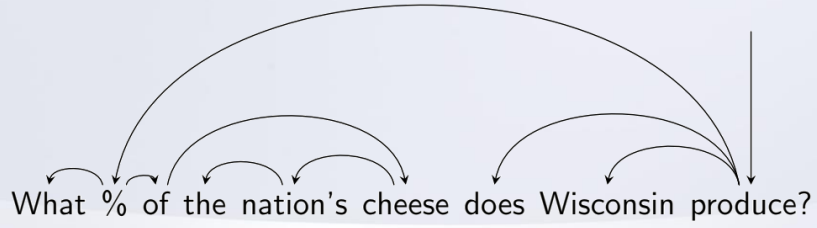
- Sentiment shifters are expressions that change orientation of sentiment (e.g. positive to negative or vice-versa)
  - Example: don't → I don't like this (positive to negative)
    less → With medication I feel less stress (negative to positive)
  - Also to be covered in further detail in later lecture

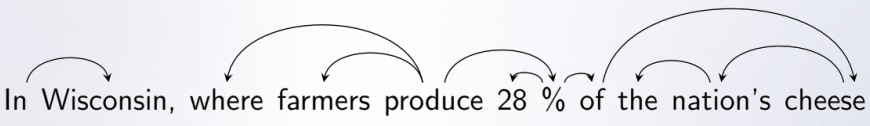
Dependencies

- Dependencies refer to the reliance of items to each other – for sentence operations there are usually semantic, morphological, prosodic and syntactic (veeeeery linguistic-oriented and out of scope for us)
- For us, syntactic dependency refers to the relation between two words in a sentence with one word being the governor and the other being the dependent → easily the most influential in terms of sentiment
  - Helps identify statements, the components, their interdependencies etc.



- Information from dependencies allow responses to be generated
  - Example before is a question, so the answer has to satisfy 'percentage' from 'Wisconsin' in the action of 'cheese production'





Source: https://github.com/jacobeisenstein/gt-nlp-class/tree/master/notes

Supervised Learning Classification

- For classification, essentially any supervised learning algorithm can be used for training a classifier; popular choices include
  - Support vector machines (SVM)
  - Logistic regression
  - Naïve Bayes methods
  - Lazy learning methods (kNN)
  - Etc.
- Choice of classifiers depend on metrics such as effectiveness, true/false positive/negative comparisons, accuracy,

- With combination of various features, a document can be classified depending on the application it is applied to, examples
  - Gamon (2004) chose features of POS trigrams, sentence length, noun phrasing and constituent structures to perform customer feedback classification
  - Mullen and Collier (2004) introduced a set of sophisticated features to combine with n-grams including PMI, opinion words, and values of adjectives
  - .....
- This course does not focus on the effort of the research/researchers but they are available for further reading in the reference textbook from last week's slides.

Related Papers

- Bickerstaffe and Zukerman (2010) used a hierarchical multi-classifier considering inter-class similarity
- Burfoot, Bird and Baldwin (2011) sentiment-classified congressional floor debates
- Cui et al. (2006) evaluated some sentiment classification algorithms
- Das and Chen (2001) extracted market sentiment from stock message boards
- Dasgupta and Ng (2009) used semi-supervised learning
- Dave, Lawrence & Pennock (2003) designed a custom function for classification
- Gamon (2004) classified customer feedback data

- Goldberg and Zhu (2006) used semi-supervised learning.
- Kim, Li and Lee (2009) and Paltoglou and Thelwall (2010) studied different IR term weighting schemes
- Li et al (2010) made use of different polarity shifting.
- Li, Huang, Zhou and Lee (2010) used personal (I, we) and impersonal (they, it, this product) sentences to help
- Maas et al (2011) used word vectors which are latent aspects of the words.
- Mullen and Collier (2004) used PMI, syntactic relations and other attributes with SVM.
- Nakagawa, Inui and Kurohashi (2010) used dependency relations and CRF.

Custom Score Function Classification

- Instead of using a standard machine learning method, researchers have proposed customized techniques specifically for sentiment classification
- One example is score technique (Dave et al 2003) which makes use of identified positive and negative words in a review document in two steps
  - Step 1: score each term (uni/n-gram) in training set with equation

$$\text{score}(t_i) = \frac{\Pr(t_i|C) - \Pr(t_i|C')}{\Pr(t_i|C) + \Pr(t_i|C')},$$

where t_i is the term, C is class, C' is the complement thus $\Pr(t_i|C)$ is probability of term t_i in C

  - Step 2: Classify a new document $d_i = t_i \ldots t_n$ by summing up the scores of all terms and using the sign of the total to determine the class

$$\text{class}(d_i) = \begin{cases} C & \text{eval}(d_i) > 0 \\ C' & \text{otherwise} \end{cases} \quad \text{where } \text{eval}(d_i) = \sum_j \text{score}(t_j)$$

- Results of experimentation show that bigrams and trigrams give the best accuracies and no stemming or stop word removal is done.

UNSUPERVISED LEARNING METHODS

Unsupervised Sentiment Classification

- Because sentiment words and phrases are often the dominating factor for sentiment classification, it is not hard to imagine using them for sentiment classification in an unsupervised manner.
- Two methods to be covered
  - One, based on the method in Turney (2002), performs classification using some fixed syntactic patterns that are likely to express opinions.
  - The other is based on a sentiment lexicon, which is a list of positive and negative sentiment words and phrases

Syntactic Pattern Matching

- In this approach, each syntactic pattern is a sequence of POS tags alongside some constraints

	First word	Second word	Third word (not extracted)
1	JJ	NN or NNS	anything
2	RB, RBR, or RBS	JJ	not NN nor NNS
3	JJ	JJ	not NN nor NNS
4	NN or NNS	JJ	not NN nor NNS
5	RB, RBR, or RBS	VB, VBD, VBN, or VBG	anything

e.g. JJ + NN/NNS = adjective + noun(s)

Syntactic Pattern Example

- Tunney (2002) approach takes three steps
  - Step 1: Extract words that match the POS patterns. Example "This piano makes beautiful sounds" will have two words "beautiful sounds" match pattern 1.
  - The reason these patterns are used is that JJ, RB, RBR, and RBS words often express opinions or sentiments. The nouns or verbs act as the contexts because in different contexts, a JJ, RB, RBR, and RBS word may express different sentiments.

- Step 2: Each identified string of patterns are calculated for their sentiment orientation using PMI (pointwise mutual information) which measures the statistical dependence between terms

$$PMI(term_1, term_2) = log_2\left(\frac{\Pr(term_1 \wedge term_2)}{\Pr(term_1)\Pr(term_2)}\right)$$

- The SO of a phrase is computed based on its association with words excellent (positive) and poor (negative)

$$SO(phrase) = PMI(phrase, \text{"excellent"}) - PMI(phrase, \text{"poor"})$$

- Step 3: the average SO score of all phrases in the document is computed, and the review is classified as positive if the average SO value is positive and negative otherwise

Sentiment Lexicon Classification

- The lexicon-based unsupervised classification performs classification based on a dictionary of sentiment words and phrases, called a sentiment/opinion lexicon → Example lexicons include SocialSent from Stanford, VADER, TextBlob and Sentiwordnet
- In its base form, to classify a document, the SO values of all sentiment expressions in the document are summed up → if overall sum is positive, document sentiment is positive etc.
  - There are many approaches to the calculation, each with pros and cons

Example

- Consider this problem instance: "Sam is a great guy."
- Tokenization is performed

"Sam is a great guy."

Tokenize
Sam
is
a
great
guy
.

- Operations for stop word removal, lemmatization, stemming etc. → followed by identifying individual sentiment

Tokenize	Preprocessing
Sam	
is	stop word
a	stop word
great	
guy	
.	punctuation

Sam	neutral
great	positive
guy	neutral

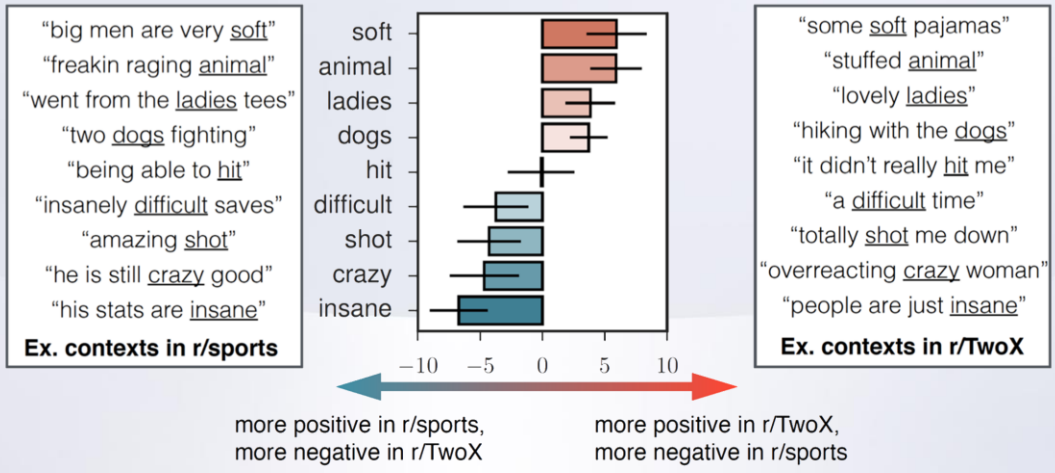
Example source: https://alphabold.com/sentiment-analysis-the-lexicon-based-approach/

- Running the lexicon on the preprocessed data, returns a **positive sentiment** score/ measurement because of the presence of a positive word "great" in the input data.

Sam	neutral
great	positive
guy	neutral

Sentiment: positive

- Each entry in the lexicon is mapped to a relevant orientation and strength → will be domain specific and needs to be chosen carefully



Source: https://nlp.stanford.edu/projects/socialsent/

Cross-Domain Classification

- Sentiment classification is highly sensitive to the domain from which the training data are extracted → mostly because the same word may be positive in one domain but negative in another
- Many approaches have been attempted with varying success including
  - training on a mixture of labeled reviews from other domains where such data are available and testing on the target domain
  - training a classifier, but limiting the set of features only to those observed in the new target domain
  - using ensembles of classifiers from domains with available labeled data and testing on the target domain
  - combining small amounts of labeled data with large amounts of unlabeled data in the target domain (this is the traditional semi-supervised learning setting).

Cross-Language Classification

- Majority of research, approaches and algorithms focus on English as the language and the others are left behind (esp. non-Romanized ones)
- Similar to cross-domain, there are many approaches to classification including
  - Retraining classifiers for specific languages (obvious approach but introduces its own problems)
  - Processing/Translation back to English to perform classification
  - Co-training SVM to learn two/more classifiers for simultaneous classification

Summary

- Content covered today
  - Document level sentiment analysis
  - Supervised / unsupervised classification