Opinion Mining

Question Answering Seminar

January 20, 2012

Nils Rethmeier
HPI Potsdam

Overview

Motivation
- Applications and the task at hand

Introduction
- Opinion definition
- Opinion analysis
  - sentences, documents, results
- Backgrounds (Bayes Classification)
- Detection features

Evaluation
- Testsets
  - documents, sentences
- Results

Discussion

Application areas

Information extraction — discard subjective results
- bias in news

Question Answering — opinion detection

Summarization — summarizing different points of view

Content rating — via comments, stars
- child protection
- appropriate ad placement

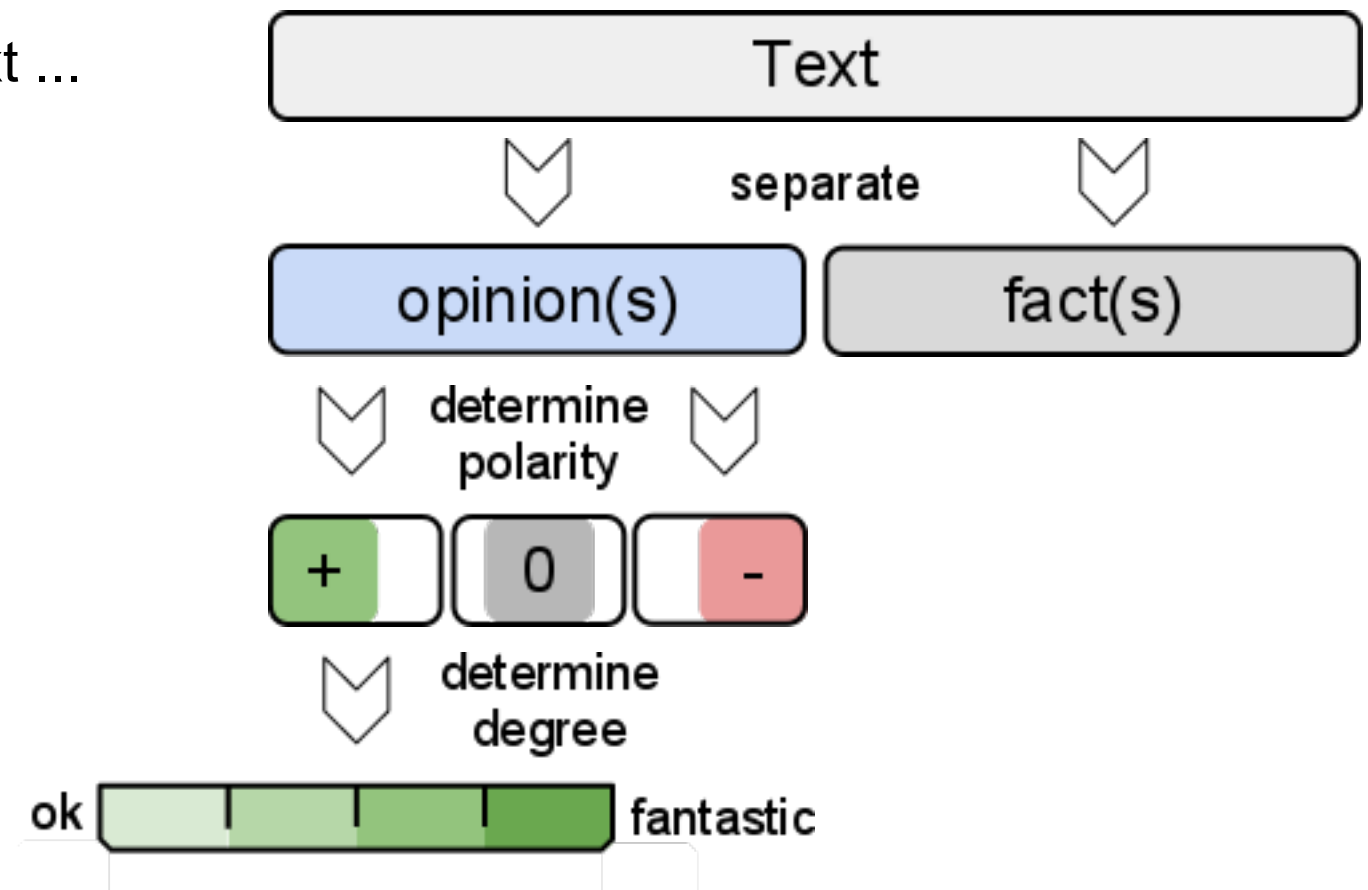
Business Intelligence — customer support
- product image mining
- help customers find needed information

Introduction

Definition

Opinion :=

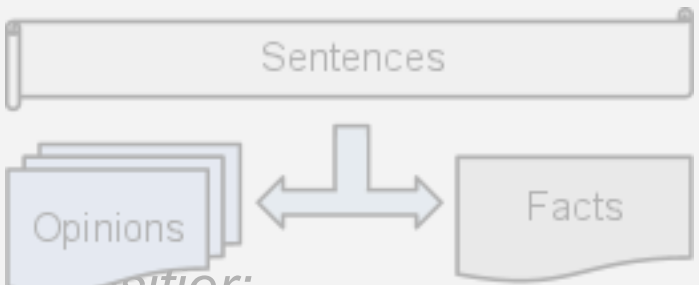
Task: Given a text ...



Classification

Sentence-level classification

Hypothesis: Opinion documents mostly contain opinion sentences

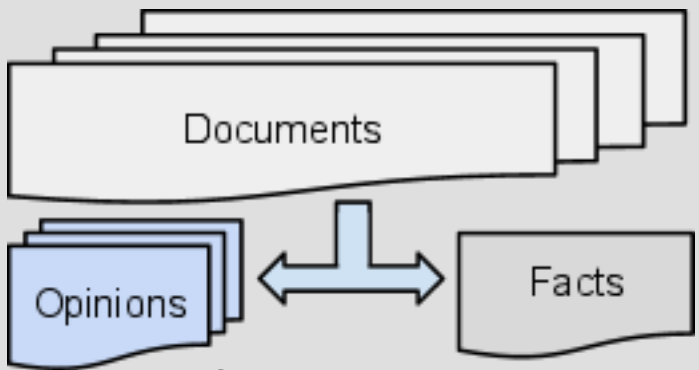


Classifier:

- sentences similarity
- 1 or n Naive Bayes

Polarity Classification

Document-level classification



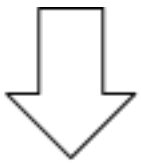
- *Classifier:* Naive Bayes
- *Training Data:* Reference text collections = News, Business articles (facts), editorials and letters to author (opinion)

Bayes Classification, theorem

Conditional Probability

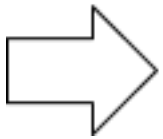
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability of A if B is known.



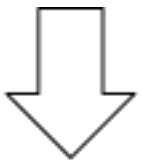
Reversed condition

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

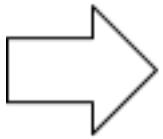


Multiplication Anxiom

$$P(A \cap B) = P(A|B)P(B)$$



Bayes' Theorem



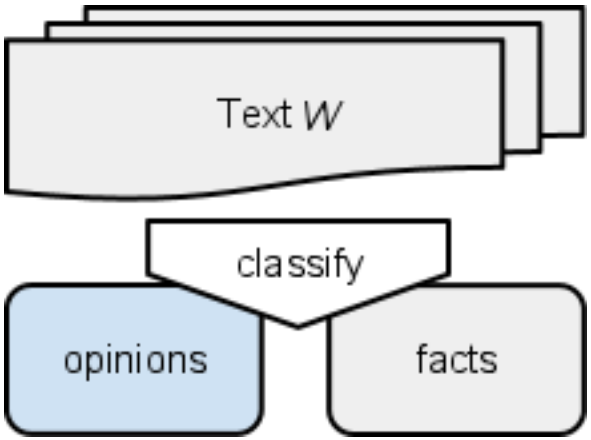
$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Bayes' Classification, steps

Bayes' Classifier (machine learning ML)

Given: Text W, of words w_i

$$W = w_1 w_2 \dots w_n$$
$$w_i := word$$



Task: Classify whether W is *opinion* or *fact*?

H $P(opinion|W) > P(fact|W)$ ...

Problem: undetermined probabilities

$$P(opinion|W) = \frac{P(W|opinion)P(opinion)}{P(W)}$$

How likely is a text if we know its an opinion?

$$P(fact|W) = \frac{P(W|fact)P(fact)}{P(W)}$$

How likely is a text if we know its a fact?

Bayes' Classification, steps

Bayes' Classifier (machine learning ML)

Problem:

$$P(opinion|W) = \frac{P(W|opinion)P(opinion)}{P(W)}$$

Though given text W $\neq$ Reference text W, we assume that Reference statistics are equal for all text.

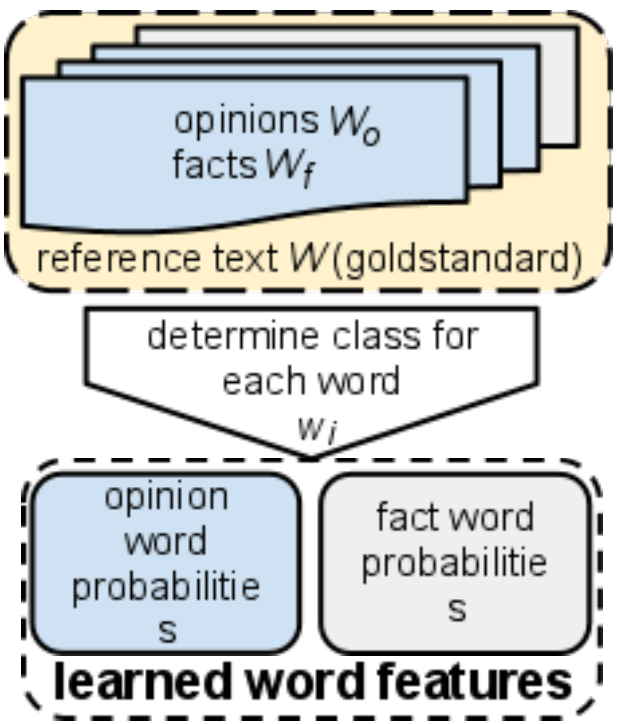
Solution:

- Take a set of reference opinions and facts
- Assume, words occur independent (Naive Bayes Assumption NBA)

$$P(W|opinion) \overset{NBA}{=} P(w_1|opinion) \dots P(w_n|opinion)$$

$$P(w_i|opinion) = \frac{\text{Number of opinion texts } W_o \text{ containing } w_i}{\text{Number of all opinion texts } W_o}$$

$$P(opinion) = \frac{\text{Number all of opinion text } W_o}{\text{Number of all reference texts } W}$$



Bayes' Classification, steps

Bayes' Classifier (machine learning ML)

Summary:

1. Learn features

   How likely is a text ***W*** **given** we want ***opinions***?

$$P(W|opinion) \overset{NBA}{=} P(w_1|opinion)\dots P(w_n|opinion)$$

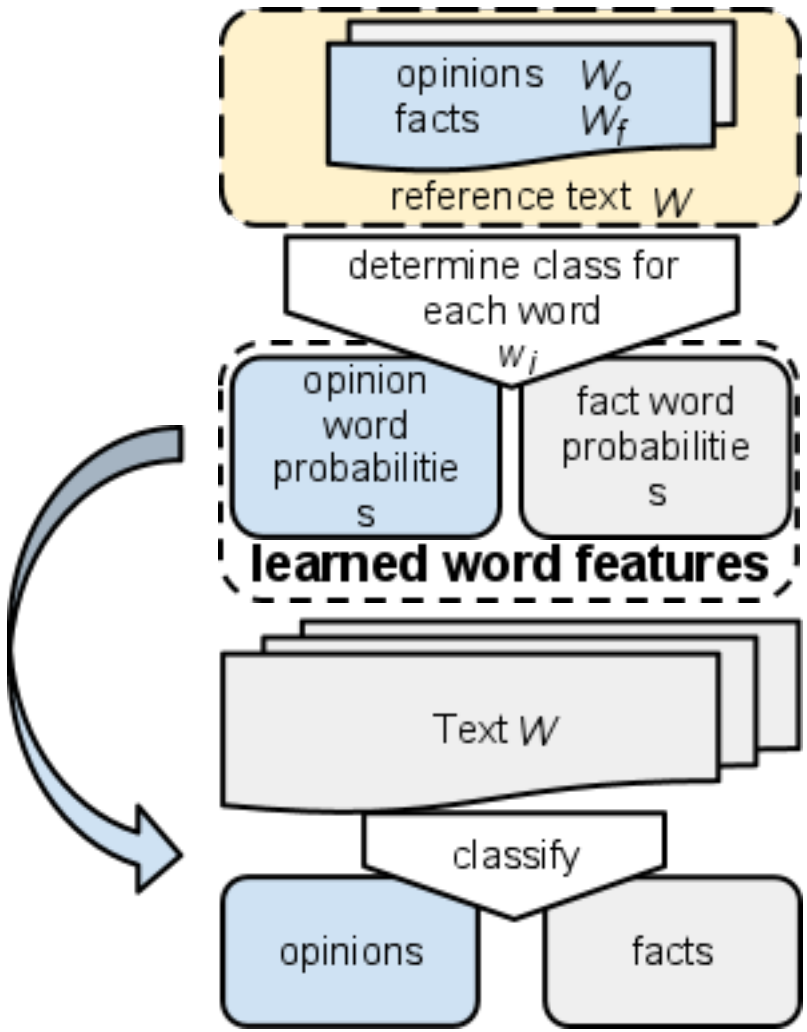
$$P(opinion) = \frac{\text{Number all of opinion text } R_o}{\text{Number all of text } R}$$

2. Use features to classify using Bayes

   How likely is an ***opinion/fact*** **given** a text ***W***?

$$P(opinion|W) = \frac{P(W|opinion)P(opinion)}{\cancel{P(W)}} \quad >$$

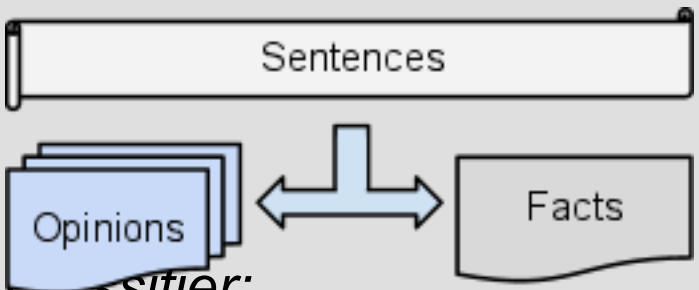
$$P(fact|W) = \frac{P(W|fact)P(fact)}{\cancel{P(W)}}$$



Classification

Sentence-level classification

Hypothesis: Opinion documents mostly contain opinion sentences

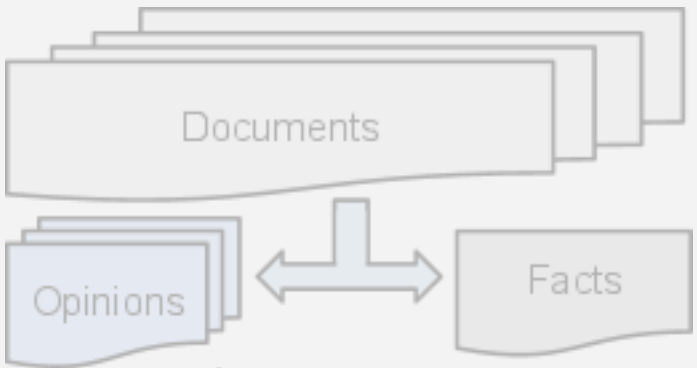


Classifier:

- sentences similarity
- 1 or n Naive Bayes

Polarity Classification

Document-level classification



- *Classifier:* Naive Bayes
- *Training Data:* Reference text collections = News, Business articles (facts), editorials and letters to author (opinion)

Classifiers: SimFinder

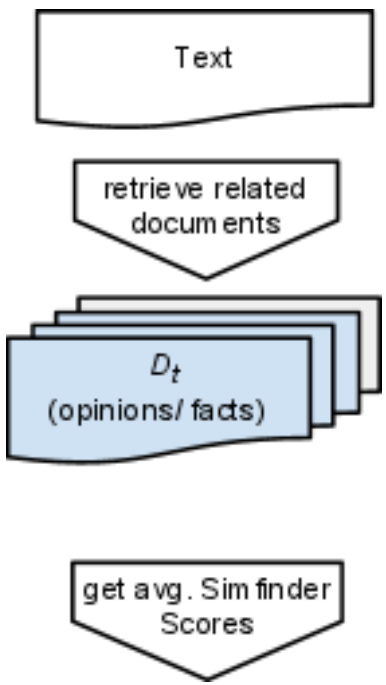
Sentence Similarity:

Idea: Given *a fixed topic*, opinion sentences are more similar to each other than they are to factual sentences.

Retrieve: All documents D_t for a topic, e.g. "welfare reforms"

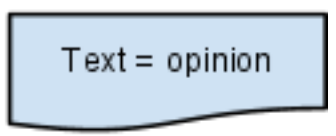
Features: SimFinder similarity score S of each sentence in D_t

- words
- phrases (n-grams)
- WordNet synsets



Classification:

$$S = \frac{S_o}{S_f} = \frac{\text{average opinion sentence score}}{\text{average fact sentence score}} \begin{cases} opinion & \text{if } S > 1.0 \\ fact & \text{if } S \leq 1.0 \end{cases}$$



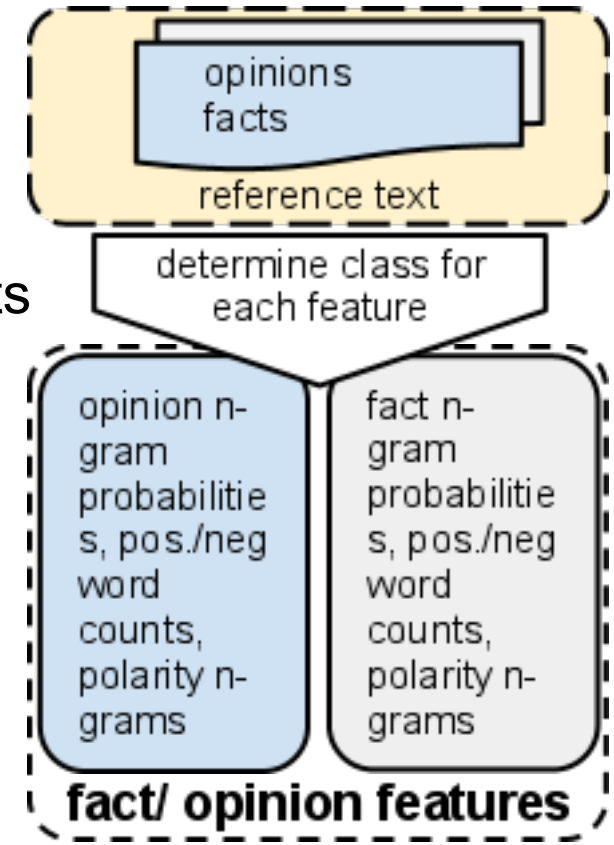
Classifier: Naive Bayes 1

1 NB classifier C on sentences

Train: Learn features on opinion/ fact articles.

Features: A classifier C with all the features

- n-grams, parts of speech (POS)
- sentence positive/ negative word counts
- polarity n-gram magnitude, e.g. "++"for two consecutive positive words



Combination:

$$P(opinion|W) = P(W|opinion)P(opinion),$$
$$\text{let } W = n\text{-}gram, POS \wedge opinion := op$$

$$P(op|n\text{-}gram, POS) = \frac{P(op|n\text{-}gram)P(op|POS)}{P(op)}$$

Classifier: Naive Bayes n

HPI

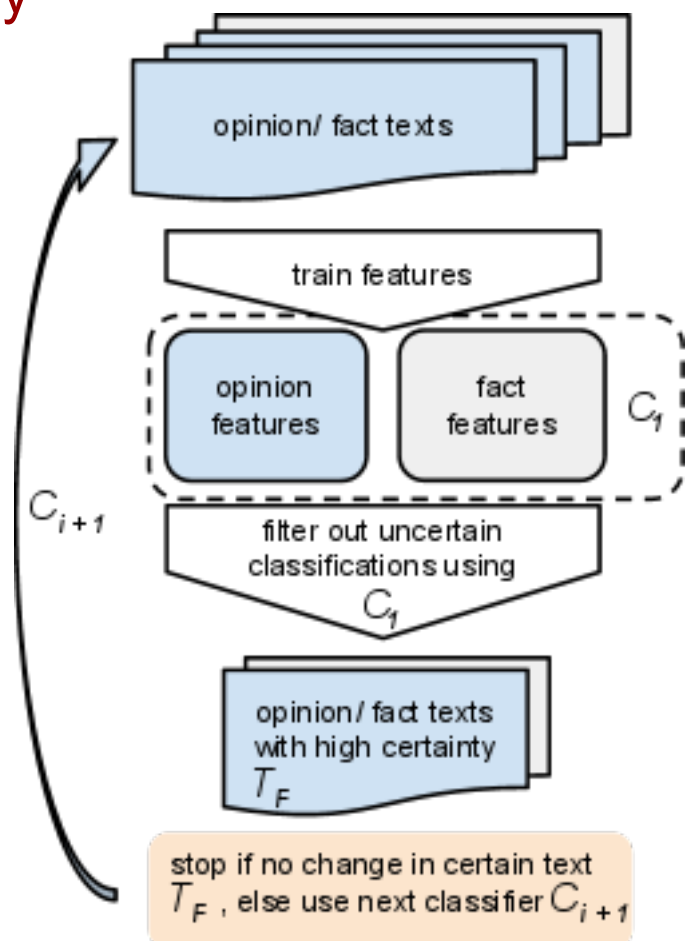
n NB classifiers C_1 .. Cn, each with a different feature

Problem: The hypothesis, that opinion documents only contain opinion sentences is flawed.

Idea: Now, only use sentences that are likely to be labeled correctly during training.

Features: as before, but split between classifiers C_i

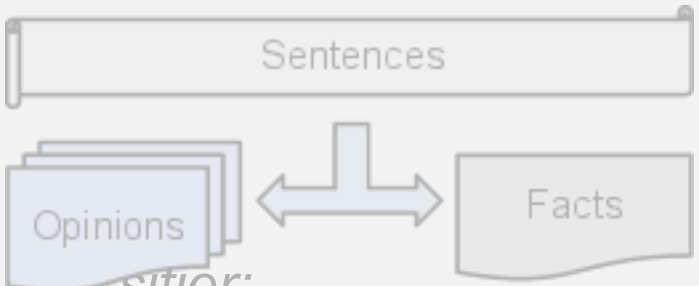
- 1-3 grams | POS | +/-words | magnitudes
- recursive filtering of the training data using next C_i at each recursion step



Polarity Classification

Sentence-level classification

Hypothesis: Opinion documents mostly contain opinion sentences

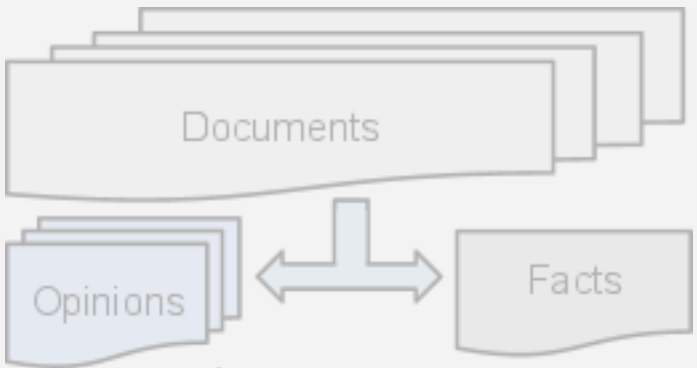


Classifier:

- sentences similarity
- 1 or n Naive Bayes

Polarity Classification

Document-level classification



- *Classifier:* Naive Bayes
- *Training Data:* Reference text collections = News, Business articles (facts), editorials and letters to author (opinion)

Polarity Classification

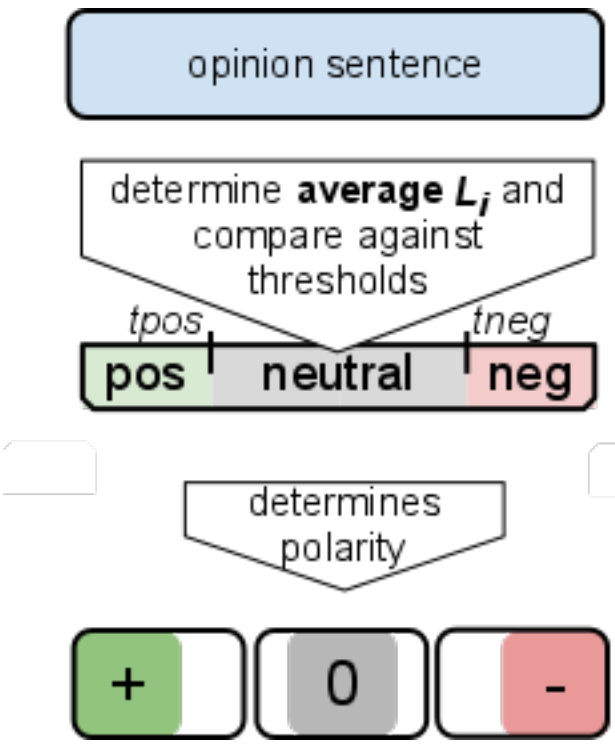
Given: A set of polarity words (manually annotated).

Idea: Positive words occur together more often than by chance (word co-occurrence).

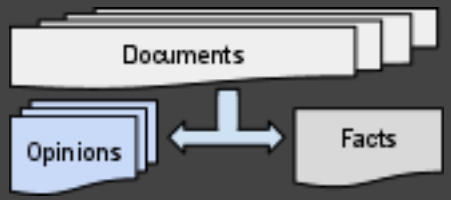
Classifier: is positive model $P(+)$ more likely?

$$L_i(w_i, POS_k) = \frac{P(+)}{P(-)} = log \left(\frac{\frac{Freq(w_i, POS_k, W_+) + \epsilon}{Freq(W, POS_k, W_+)}}{\frac{Freq(w_i, POS_k, W_-) + \epsilon}{Freq(W, POS_k, W_-)}} \right)$$

w_i := *i-th word in sentence*
POS_k = *part of speech* : k = *adj, adverb, noun, verb*
W_+ := *set of positive words*
W_- := *set of negative words*
ϵ := *smoothing constant, e.g.* $\epsilon = 0.5$



Evaluation



Trainingset: 2000 Wall Street Journal (WSJ) articles for each (=4000)
- facts from labels "news", "business articles"
- opinions from labels "editorial" and "Letter to editor"

Testset: another 2000 WSJ articles each

Documents classification

Goldstandard: label of each article

Naive Bayes classifier:

	F-measure
News vs. *Editorial*	0.96
News+Business vs. *Editorial+Letter*	0.97

Sentence classification

400 sentences of human annotations
- A=300 one annotator
- B=100 two annotators agree on type

Similarity classifier: {recall, precision}

Variant	Class	Standard A	Standard B
Score	Fact	{0.61,0.34}	{1.00,0.27}
	Opinion	{0.30,0.49}	{0.16,0.64}

Evaluation

Sentence classification

1 and n Naive Bayes classifiers: human annotations (A = 300, B = 100)

Features	Class	Standard A		Standard B	
		Single	Multiple	Single	Multiple
Words only	Fact	{0.14,0.39}	{0.12,0.42}	{0.28,0.42}	{0.28,0.45}
	Opinion	{0.90,0.69}	{0.92,0.69}	{0.90,0.82}	{0.91,0.83}
Words, Bigrams, Trigrams, Part-of-Speech, and Polarity	Fact	{0.15,0.43}	{0.13,0.42}	{0.44,0.50}	{0.44,0.53}
	Opinion	{0.91,0.69}	{0.92,0.70}	{0.88,0.86}	{0.91,0.86}

- using words only works well already
- using word n-grams + POS + polarity works best
- using multiple-classifier-filtering increases recall

Evaluation

Sentence classification
polarity classifier: accuracy

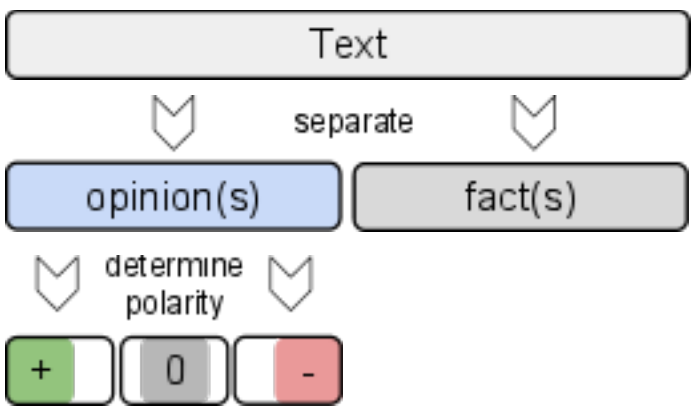
Parts-of-speech Used	A	B
Adjectives	0.49	0.55
Adverbs	0.37	0.46
Nouns	0.54	0.52
Verbs	0.54	0.52
Adjectives and Adverbs	0.55	0.84
Adjectives, Adverbs, and Verbs	0.68	0.90
Adjectives, Adverbs, Nouns, and Verbs	0.62	0.74

verbs and verbs yields
n

Discussion

Opinion Mining

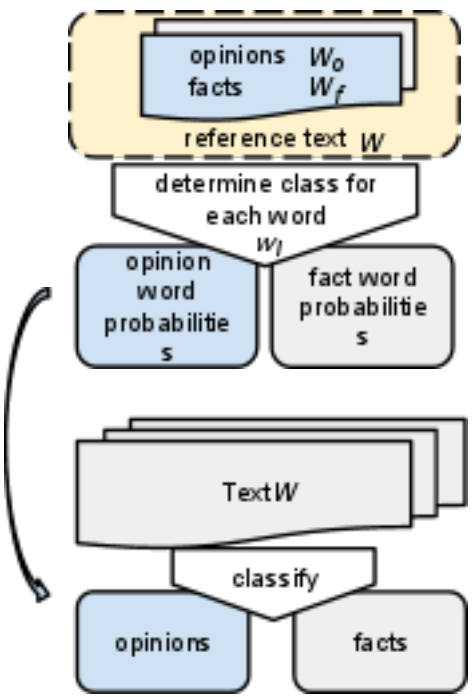
Fact/ Opinion Classification



Classifier:

- document
  - Naive Bayes
- sentences
  - similarity
  - 1 or n Naive Bayes
  - polarity

NB Classifier



Evaluation

Documents:

- Naive Bayes produces 97% F-measure

Sentences:

- Similarity less useful
- **Naive Bayes** already works well on **word n-grams** (86% precision)
- **polarity classification** needs **adjectives, adverbs and verbs** to work well (90% agreements)