SENTIMENT ANALYSIS

Maithily Pande	111203043
Pooja Kulkarni	111203074
Vitrag Shah	111208055
Smruti Charkha	111208009

Under guidance of : Y. V. Haribhakta

INTRODUCTION

Motivation

Movie Reviews : Positive or Negative

Twitter

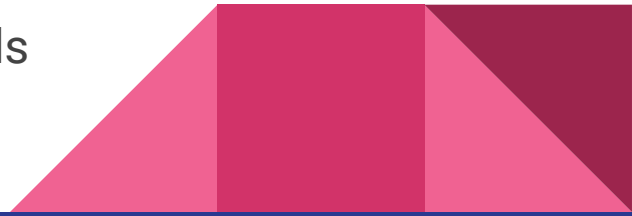

Type in a word and we'll highlight the good and the bad
"united airlines"
Search
Save this search
Sentiment analysis for "united airlines"
Sentiment by Percent
Negative (68%)
Positive (32%)
Sentiment by Count
Positive (11)
Negative (23)
0
5
10
15
20
25
30
jljacobson: OMG... Could @United airlines have 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

Other Names

- Opinion Mining
- Opinion Extraction
- Sentiment Mining
- Subjectivity Analysis

Why Sentiment Analysis?

- Product
- Public Sentiment
- Politics
- Prediction : Predict the election results, market trends



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*

Definition

Sentiment Analysis is the detection of **Attitudes**

Enduring, affectively colored beliefs, disposition towards objects or persons

1.Holder(Source) of attitude

2.Target(aspect) of attitude

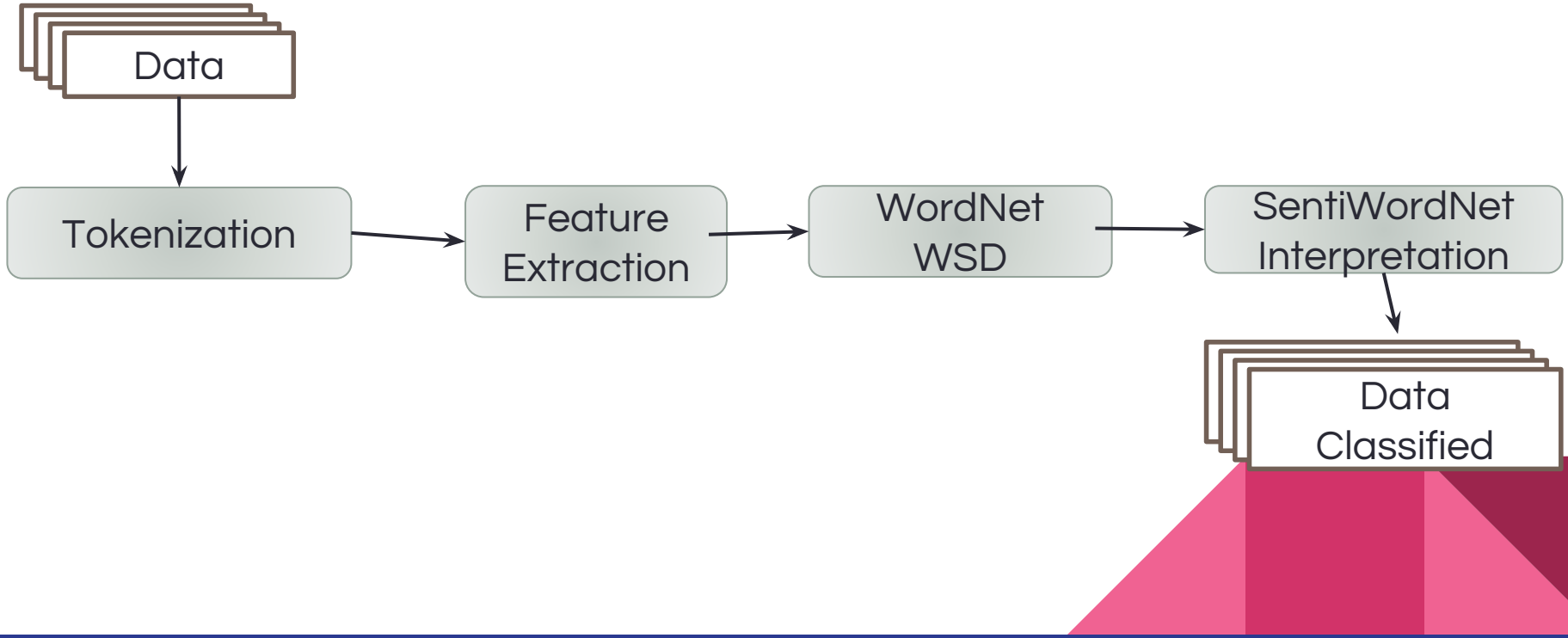
3.Type of attitude

- From a set of types: Like, Love, Hate, Value, Desire, etc
- Or (more commonly) simple weighted polarity : positive, negative, neutral

4. Text containing the attitude : sentence or entire document

HOW IT WORKS

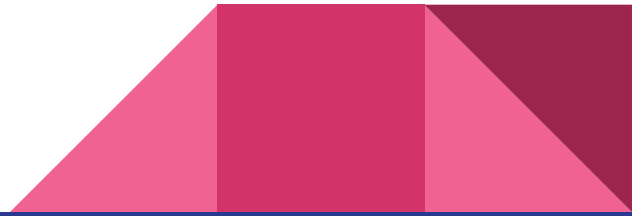
Sentiment Classification Phases



Tokenization

- **Tokenization process**:
  It splits the text into very simple tokens such as numbers, punctuation and words of different types.

  - Deals with HTML and XML markup.
  - Twitter markup (names,hashtags)
  - Capitalization(preserve for words in all caps)
  - Phone numbers,dates
  - Emoticons



Features Extraction

Speech Tagging process:

It produces a tag as an annotation based on the role of each word in the document.

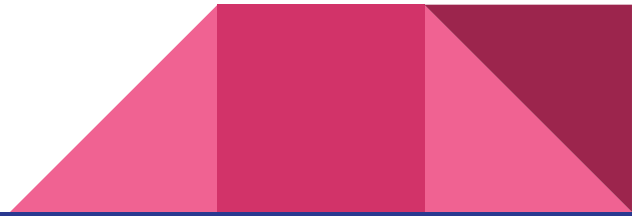
Word Sense Disambiguation :

The techniques of WSD are aimed at the determination of the meaning of every word in his context.

Extracting Features for Sentiment Classification

- How to handle negation
  - I **didn’t** like this movie
  - I really like this move

- Which words to use ?
  - Only adjectives
  - All words



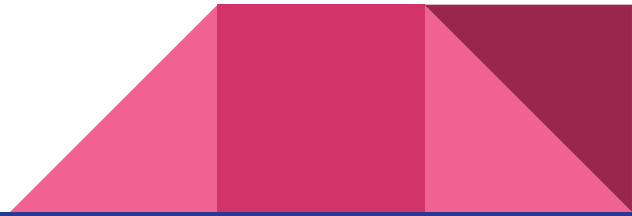
How to handle Negation ?

Add NOT_ to every word between negation and following punctuation :

For example :

1. Didn’t like this movie, but I

   Didn’t **NOT_like NOT_this NOT_movie**, but I



SentiWordNet Interpretation

Given a synset (after the phase of WSD) we can search in SentiWordNet the sentiment score associated to this synset

Data : This is very accurate and hilarious. Well done :)

Synset : **accurate#1** conforming exactly or almost exactly to fact or to a standard or performing with total accuracy; "an accurate reproduction"; "the accounting was accurate"; "accurate measurements"; "an accurate scale"

Score:

Pos_score	Neg_score	Obj_score
0.5	0	0.5

Classification Using different Classifiers

- Naive Bayes
- MaxEnt
- SVM

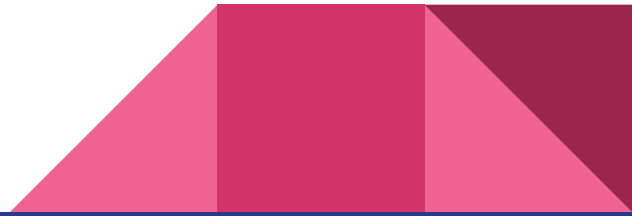
NAIVE-BAYES

HOW IT WORKS

Naive Bayes is a supervised way of sentiment analysis.

In Naive Bayes, we will have a data-set and we want to classify the set into sentiment classes - eg POS/NEG or rating 1-5

We run the Naive Bayes Training Algorithm on a part of data and have our classifier learn this data. We then predict the classes of rest of the data using this classifier.



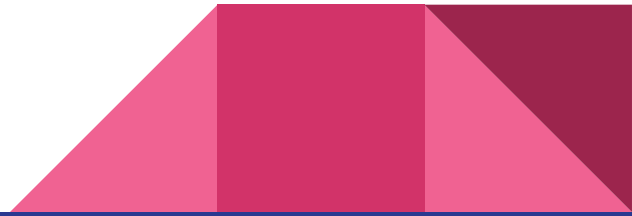
NOTATION

C_i : Any category in given categories (eg. POS, NEG for binary classification)

V : Vocabulary

W_j : Any word in the sentence

S : Any sentence of words from the vocabulary



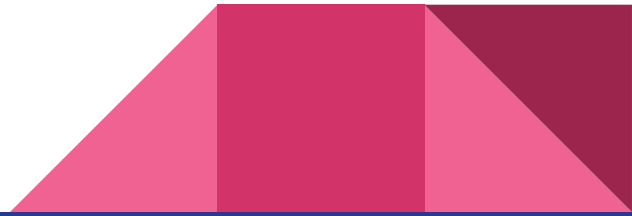
NAIVE BAYES

AIM: To find $P(C_i/S)$

We use Bayes Algorithm to find this.

$P(C_i/S) = \text{argmax}_{Ci} P(C_i) \prod_j P(W_j/C_i)$

P(w/c) = (count(w, c) + 1) / (count(c) + |V|)



EXAMPLE

TYPE	DOC	WORDS	CLASS
TRAINING	d1	Chinese Beijing Chinese	c
	d2	Chinese Chinese Shanghai	c
	d3	Chinese Macao	c
	d4	Tokyo Japan Chinese	j
TEST	d5	Chinese Chinese Chinese Toyo Japan	?

SOLUTION

|V| = 6 (Chinese, Beijing, Shanghai, Macao, Tokyo, Japan)
P(c) = 0.75
P(j) = 0.25

P(Chinese/c) = (5 + 1) / (8 + 6) = 0.4286

P(Tokyo/c) = (0 + 1) / (8 + 6) = 0.0714

P(Japan/c) = (0 + 1) / (8 + 6) = 0.0714

$P(c/d5) = P(c) * P(Chinese/c)^3 * P(Tokyo/c) * P(Japan/c)$

= 0.000301

P(Chinese/j) = (1 + 1) / (3 + 6) = 0.22

P(Tokyo/j) = (1 + 1) / (3 + 6) = 0.22

P(Japan/j) = (1 + 1) / (3 + 6) = 0.22

$P(j/d5) = P(c) * P(Chinese/c)^3 * P(Tokyo/c) * P(Japan/c)$

= 0.000128

DEMO

CHALLENGES

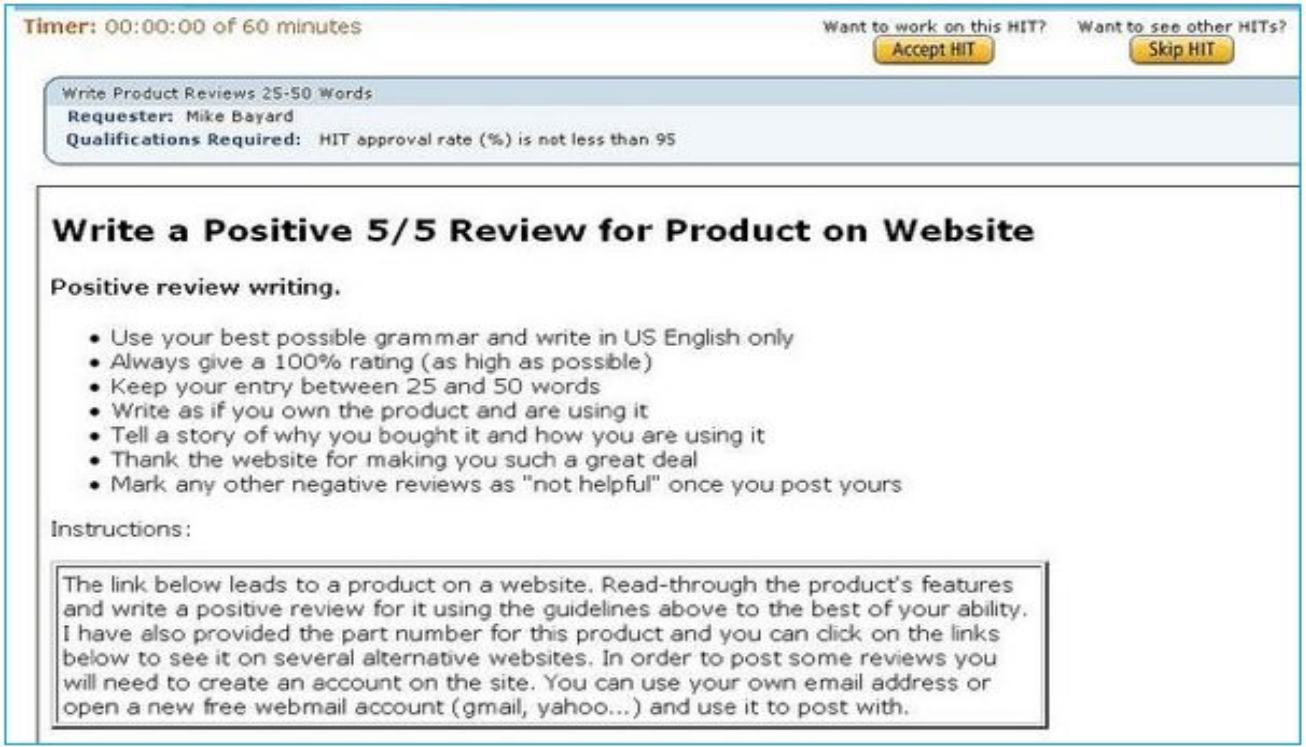
- General way of representation:

  > (O, F, s, h, t)

  > (O1, O2, F, po, h, t)

- Fake reviews.

Challenges


Timer: 00:00:00 of 60 minutes
Want to work on this HIT?
Accept HIT
Want to see other HITs?
Skip HIT
Write Product Reviews 25-50 Words
Requester: Mike Bayard
Qualifications Required: HIT approval rate (%) is not less than 95
Write a Positive 5/5 Review for Product on Website
Positive review writing.
• Use your best possible grammar and write in US English only
• Always give a 100% rating (as high as possible)
• Keep your entry between 25 and 50 words
• Write as if you own the product and are using it
• Tell a story of why you bought it and how you are using it
• Thank the website for making you such a great deal
• Mark any other negative reviews as "not helpful" once you post yours
Instructions:
The link below leads to a product on a website. Read-through the product's features and write a positive review for it using the guidelines above to the best of your ability. I have also provided the part number for this product and you can click on the links below to see it on several alternative websites. In order to post some reviews you will need to create an account on the site. You can use your own email address or open a new free webmail account (gmail, yahoo...) and use it to post with.

THANK YOU!

REFERENCES:

1. https://www.dataquest.io/blog/naive-bayes-movies/
2. http://arxiv.org/pdf/1305.6143.pdf
3. https://class.coursera.org/nlp/lecture/preview
4. https://snippetsofcode.wordpress.com/2014/04/28/fast-tutorial-to-nltk-using-python/