Sentiment Analysis: Tools and Techniques

Presented By: Clique

Anuradha Vakil

Dhanvantari Tilak

Gaurangi Bhangale

Guided By:

Prof. Chandrasekar Vuppalapati

Sentiment Analysis
What People Think?

What others think has always been an important piece of information

"Which car should I buy?"

"Which schools should I apply to?

"Which company to work for?"

"Whom should I vote for?"

So whom shall I ask?

Pre Web

- *Friends and relatives*
- *Acquaintances*
- *Consumer Reports*

Post Web

"...I don't know who..but apparently it's a good phone. It has good battery life and..."

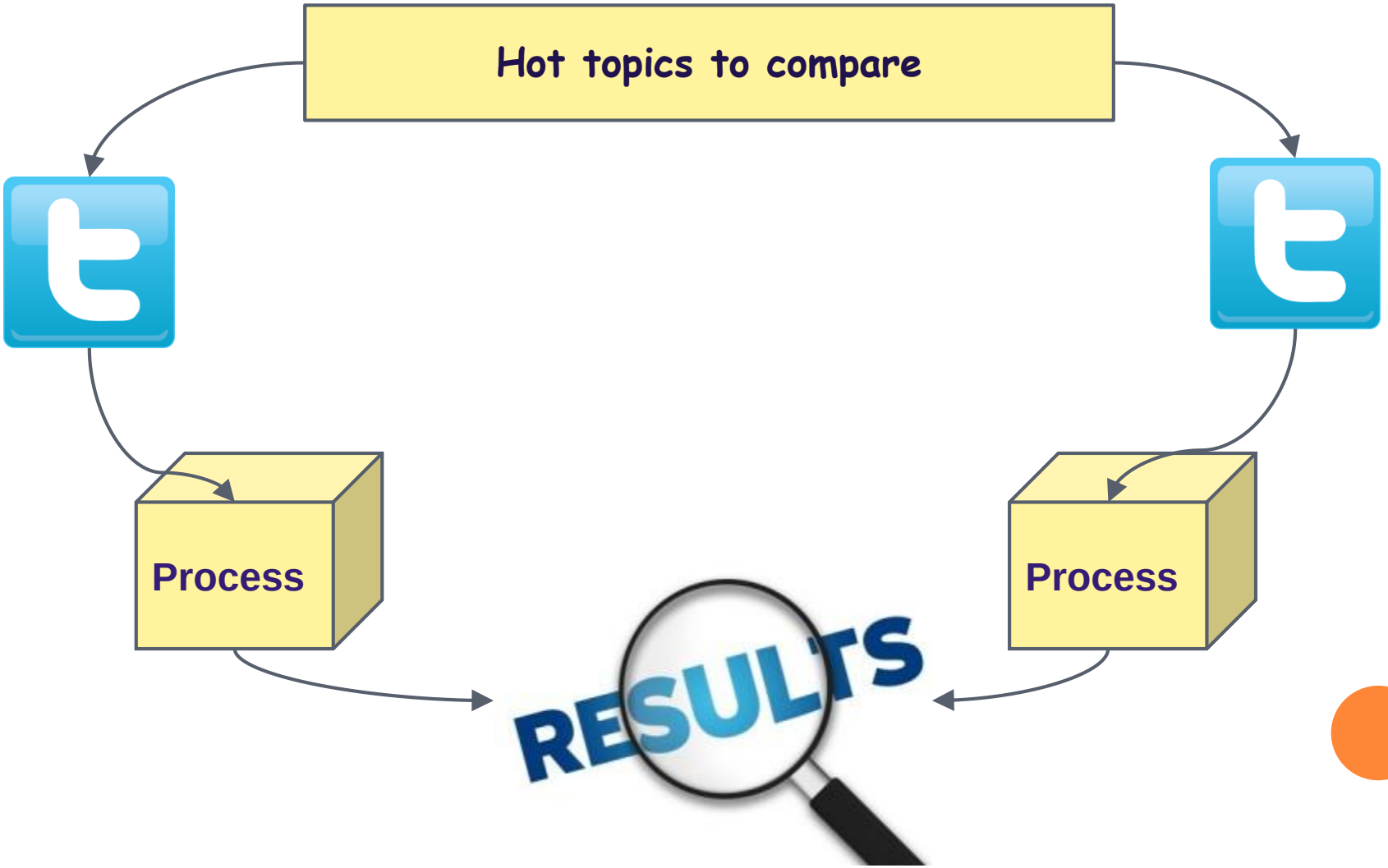
- *Blogs (google blogs, livejournal)*
- *E-commerce sites (amazon, ebay)*
- *Review sites (CNET, PC Magazine)*
- *Discussion forums (forums.craigslist.org, forums.macrumors.com)*
- *Friends and Relatives (occasionally)*

What next?

"Whoala!
I have the reviews I need"

But now I have too much of information!!
What to do with that?

Our Project



Research Work

Classifiers:

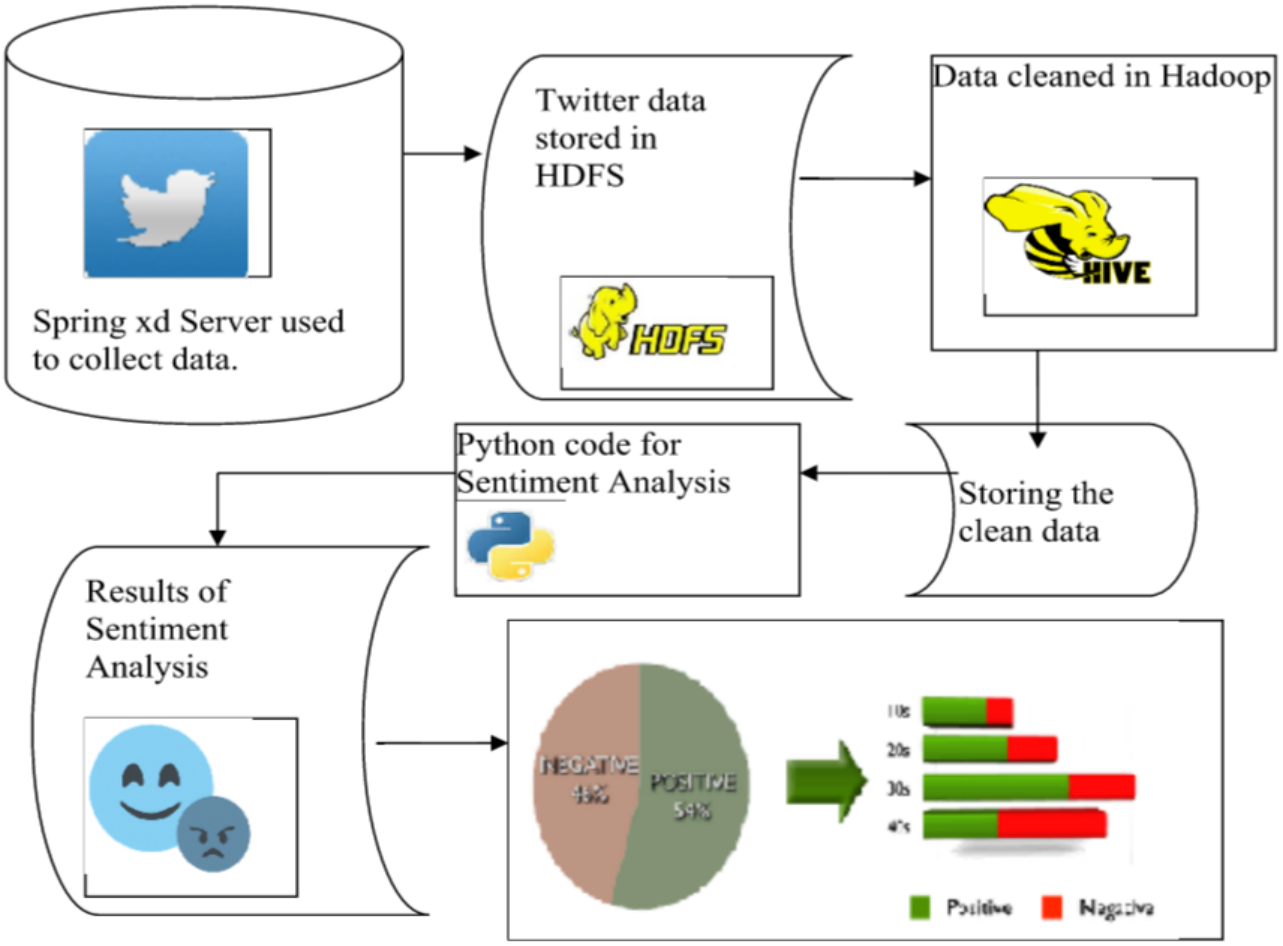
Classification	Adv	Disadv
1. Naive Bayes Classifier	Simple, Intuitive	Assumes conditional independence
2. SVM	Universal learners	Moderate results
3. Neural networks	Superior results	Slow learners
4. Decision tree classifier	Fast and scalable	Instability
5. Rule Based Classifier	Highly Expressive	Comparatively less accurate
6. Maximum Entropy Classifier	Suitable for text classification	Optimizing is comparatively slow
7. Lexicon Based Approach	Saves time	Need to maintain separate dictionary

Tools: Meltwater, Google analytics, Social mention

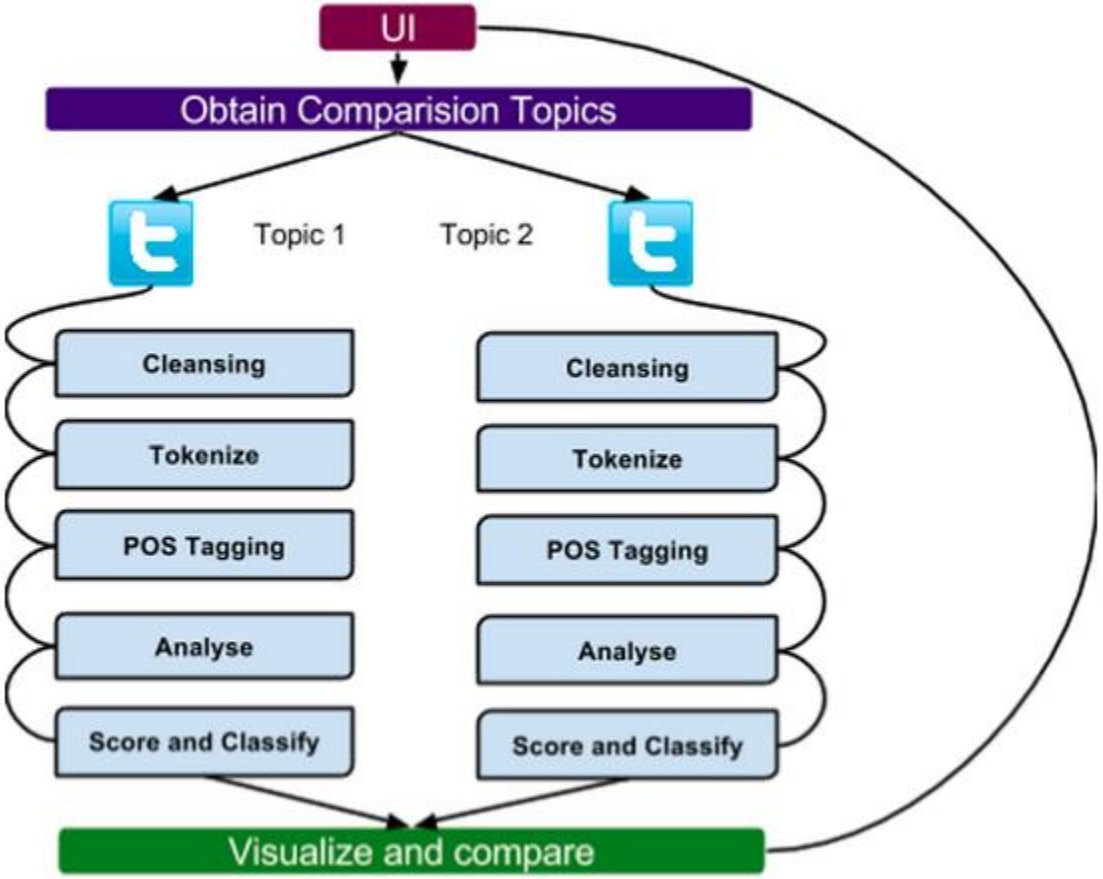
Related Fields To Sentimental Analysis

- Emotion Detection:
  - SA task
  - More into finding exact emotion
- Building Resources
  - Creating Dictionaries
  - Not exactly SA but can be used to improve SA and ED
- Transfer Learning
  - Extracting information from auxiliary domain to improvise learning in target domain

High Level Architecture



Data Flow


UI
Obtain Comparision Topics
Topic 1
Topic 2
Cleansing
Tokenize
POS Tagging
Analyse
Score and Classify
Cleansing
Tokenize
POS Tagging
Analyse
Score and Classify
Visualize and compare

Deciding on structure of our text

Lexicon based sentiment analysis--- with the help of dictionaries.
NLTK Library used for -- split tweets into sentences, tokenization and POS tagging

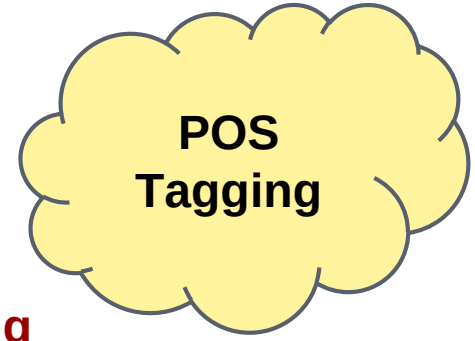
- Each tweet is a list of sentences
- Each sentence is a list of tokens
- Each token is a tuple of three elements: a word form (the exact word that appeared in the text), a word lemma (a generalized version of the

Sentence segmentation *Identify sentence boundaries*	Frank met the president. He said: "Hi! What's up – Mr. President?"	**Sentence 1:** Frank met the president. **Sentence 2:** He said: "Hi What's up – Mr. President?"
Tokenization *Identify word boundaries*	My phone tries to change 'eating' to 'dating'. #hateautocorrect	[My] [phone] [tries] [to] [change] ['] [eating] ['] [to] ['] [dating] ['] [.] [#hateautocorrect]
Stemming/lemmatization	eating, ate, eat	eat, eat, eat
Part-of-Speech tagging	If you build it, he will come	If you build it , he will come IN PRP VBP PRP , PRP MD VB

cmpe 239 is very interesting

cmpe 239 is very interesting

['cmpe', '239', 'is', 'very', 'interesting']



cmpe 239 is very interesting

['cmpe', '239', 'is', 'very', 'interesting']

[('cmpe', 'cmpe', ['NN']), ('239', '239', ['NN']), ('is', 'is', ['VBZ']), ('very', 'very', ['RB']), ('interesting', 'interesting', ['JJ'])]

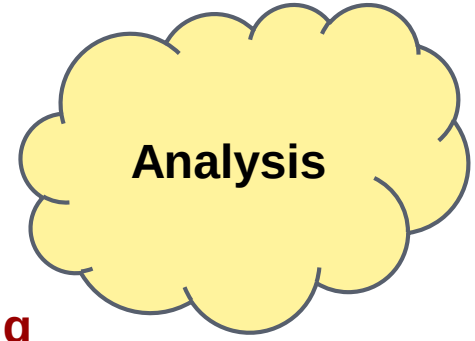
Dictionary Comparision

cmpe 239 is very interesting

['cmpe', '239', 'is', 'very', 'interesting']

[('cmpe', 'cmpe', ['NN']), ('239', '239', ['NN']), ('is', 'is', ['VBZ']), ('very', 'very', ['RB']), ('interesting', 'interesting', ['JJ'])]

[('cmpe', 'cmpe', ['NN']), ('239', '239', ['NN']), ('is', 'is', ['VBZ']), ('very', 'very', ['inc', 'RB']), ('interesting', 'interesting', ['positive', 'JJ'])]



cmpe 239 is very interesting

['cmpe', '239', 'is', 'very', 'interesting']

[('cmpe', 'cmpe', ['NN']), ('239', '239', ['NN']), ('is', 'is', ['VBZ']), ('very', 'very', ['RB']), ('interesting', 'interesting', ['JJ'])]

[('cmpe', 'cmpe', ['NN']), ('239', '239', ['NN']), ('is', 'is', ['VBZ']), ('very', 'very', ['inc', 'RB']), ('interesting', 'interesting', ['positive', 'JJ'])]

analyzing sentiment...

cmpe 239 is very interesting

['cmpe', '239', 'is', 'very', 'interesting']

[('cmpe', 'cmpe', ['NN']), ('239', '239', ['NN']), ('is', 'is', ['VBZ']), ('very', 'very', ['RB']), ('interesting', 'interesting', ['JJ'])]

[('cmpe', 'cmpe', ['NN']), ('239', '239', ['NN']), ('is', 'is', ['VBZ']), ('very', 'very', ['inc', 'RB']), ('interesting', 'interesting', ['positive', 'JJ'])]

analyzing sentiment…

Positive

Demo

http://sentitweets.appspot.com

Thank You