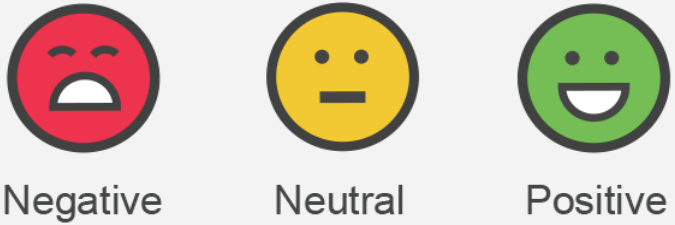
Twitter Sentiment Analysis



Presented by:

Loitongbam Gyanendro Singh

What is “Sentiment analysis”?

- Study that aims to identify the orientation of opinions in a text

Source of Sentiment Information

Source of image: http://jameskaskade.com/?p=2336

Why Sentiment Analysis?

Advent of various social media platforms

- ➔ Given netizen the liberty to openly express their views and opinions
- ➔ Large volume of data to get these information
- ➔ Knowing "what people think"
- ➔ Studies of SA deals:
  - ◆ Product and services reviews,
  - ◆ Celebrities,
  - ◆ Government policies,
  - ◆ Event,
  - ◆ and many more...

OM: Study the subjectivity of opinion
SA: Study the sentiment of opinion

Opinion

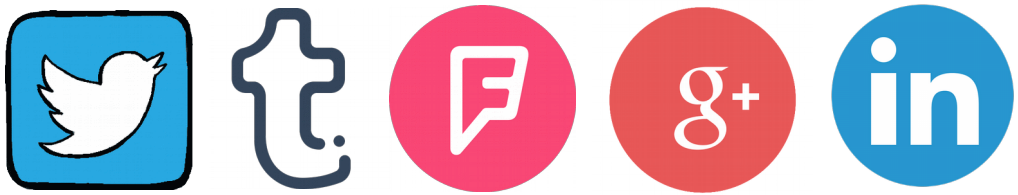
- An opinion is quintuple: (Bing Liu, 2012)
  - $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$

- Example:
  - The picture quality of my new Nikon V3 camera is great
  - (Nikon V3, picture quality, positive, User, Time)

- Where can we find opinionated text?
  - Blogs
  - Microblogs
  - Consumer forum/sites, etc.

Microblogs

- Microblogs contains a large amount of opinionated text
- There are many microblogging platforms available
  - **Twitter**
  - Tumbler
  - FourSquare
  - Google+
  - LinkedIn



- Twitter provides an easy way to access and download published posts

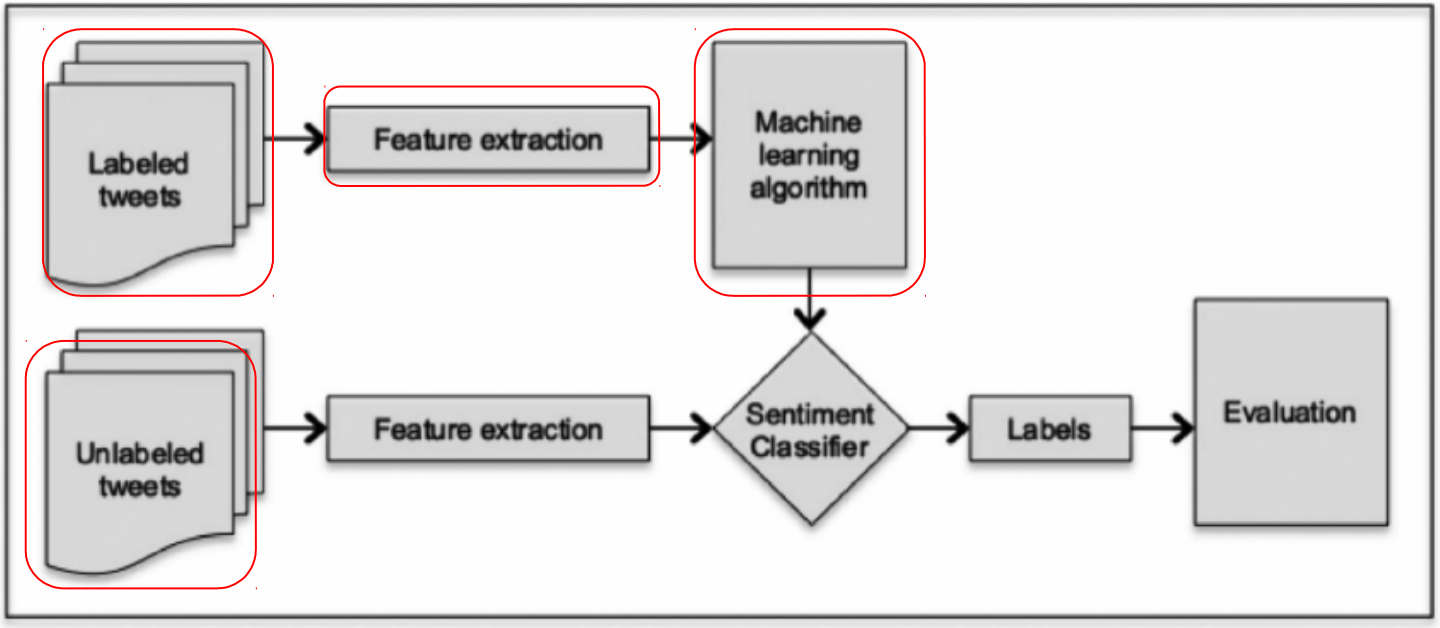
Microblogs

- ➔ Microblogs contains a large amount of opinionated text
- ➔ There are many microblogging platforms available
  - **Twitter**
  - Tumbler
  - FourSquare
  - Google+
  - LinkedIn

- ➔ Twitter provides an easy way to access and download published posts

Twitter Sentiment Analysis

➔ Majority of TSA studies deals on building sentiment classifier



TSA challenges

- ➔ Text length
- ➔ Topic relevance
- ➔ Noisy text
- ➔ Data sparsity
- ➔ Negation
- ➔ Stopwords
- ➔ Tokenization
- ➔ Multilingual content
- ➔ Multimodal content

Features

- ➔ Semantic
- ➔ Syntactic
- ➔ Stylistic
- ➔ Twitter specific features

Opinion words,
Sentiment words,
Semantic concepts,
Negation, etc.

Features

- ➔ Semantic
- ➔ Syntactic
- ➔ Stylistic
- ➔ Twitter specific features

Unigrams,
Bigrams,
N-grams,
Terms' frequencies,
POS,
Dependency tree, etc

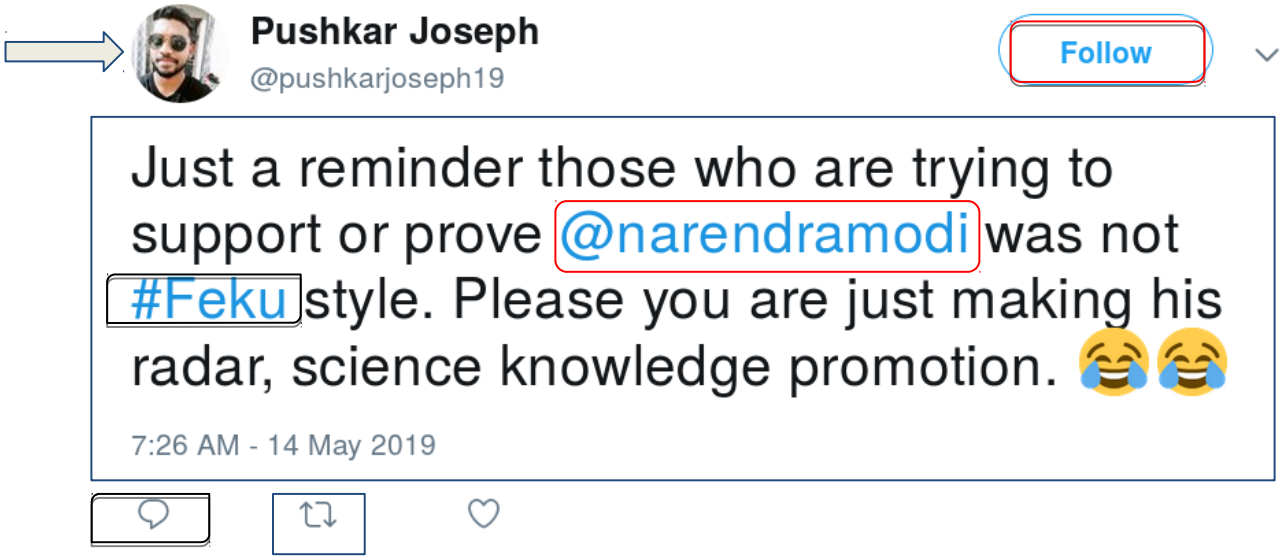
Features

- ➔ Semantic
- ➔ Syntactic
- ➔ Stylistic
- ➔ Twitter specific features

Emoticons,
Intensifiers,
Abbreviations,
Slang terms,
Punctuation marks, etc.

Twitter Specific Features

- ➔ Tweet
- ➔ User
- ➔ Mention
- ➔ Replies
- ➔ Follower
- ➔ Retweet
- ➔ Hashtag
- ➔ Privacy



Features Selection

- ➔ Manual selection
- ➔ Statistical analysis
- ➔ Dimensionality reduction
- ➔ Representation learning

Statistical Approach

- Entropy,
  - $H(X) = -\sum_{i \in C}[P(x_i) * \log(P(x_i))]$
- Strength of Association via **Pointwise Mutual Information**,
  - $PMI(x,S) = \log(P(x,S)/\{P(x)*P(S)\})$
  - $SOA(x,S) = PMI(x,S) - PMI(\neg x,S)$

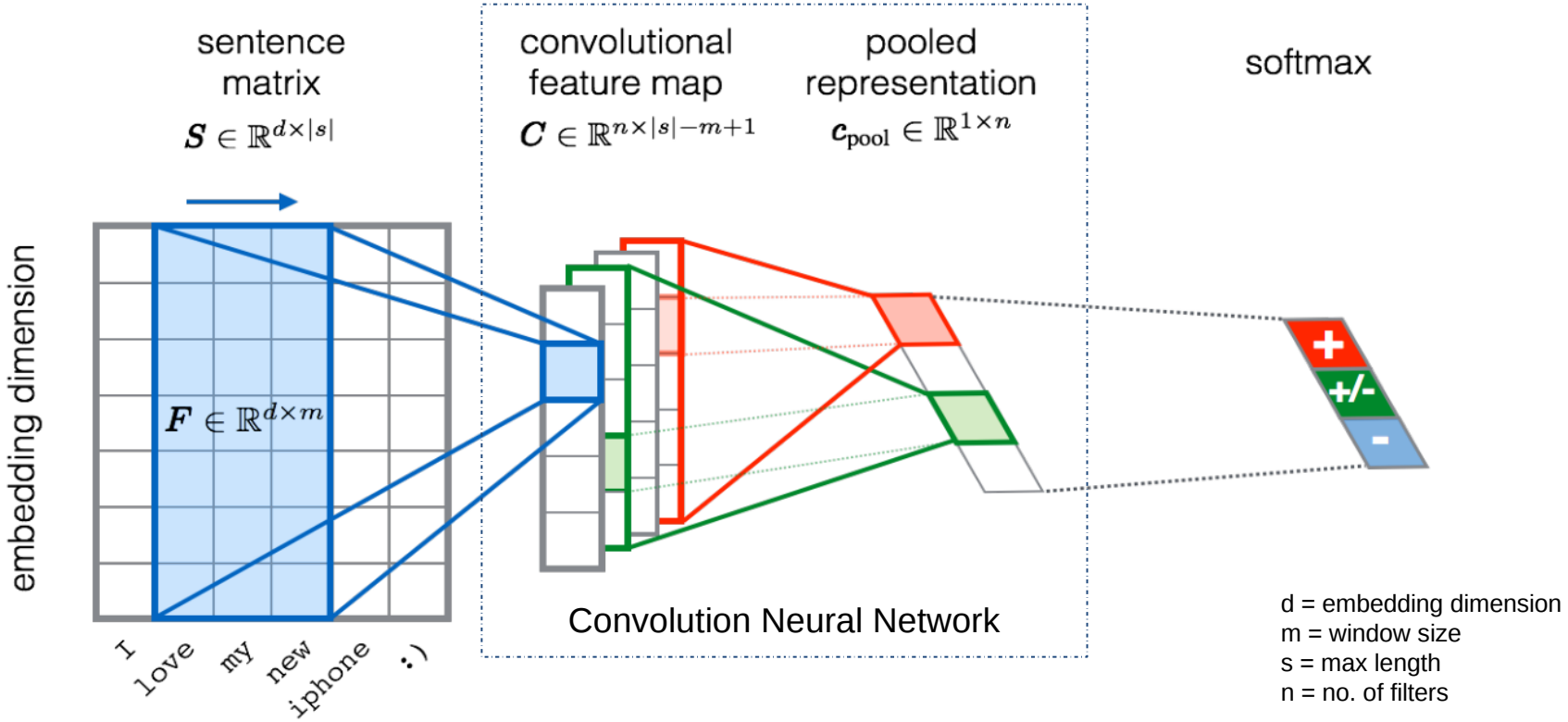
Latent Representation Methods

- Eigen Value Decomposition (EVD)
- Singular Value Decomposition (SVD)
- Word Embedding via Word2Vec, etc.

Classification Approach

- ➔ Machine Learning
- ➔ Lexicon-based
- ➔ Hybrid-based

DNN Classification approach

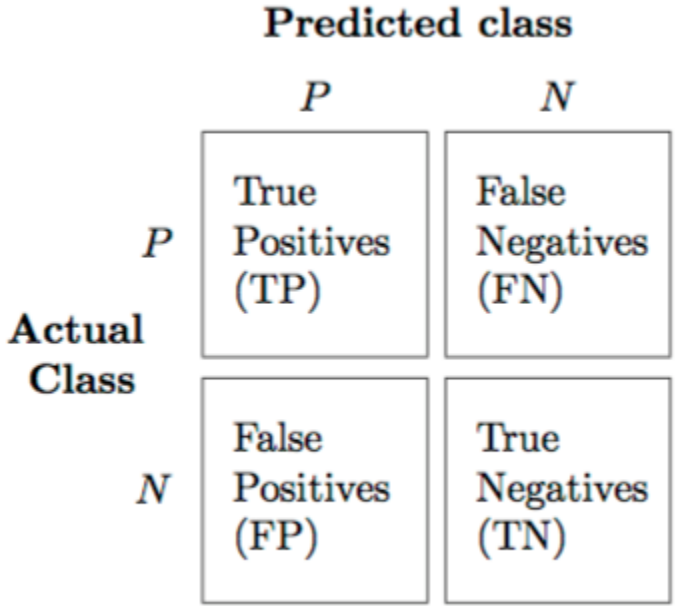


d = embedding dimension
m = window size
s = max length
n = no. of filters

Paper Title: Twitter Sentiment Analysis with Deep Convolutional Neural Networks - Aliaksei Severyn and Alessandro Moschitti

Evaluation Metrics

- ➔ Accuracy
- ➔ Precision
- ➔ Recall
- ➔ F-score



Related fields

- Twitter-based Opinion Retrieval
- Tracking Sentiment over Time
- Irony Detection on Tweets
- Emotion Detection on Tweets
- Tweet Sentiment Quantification

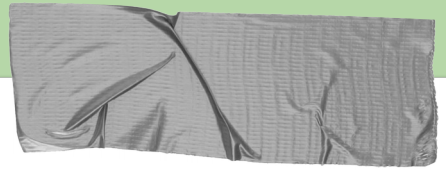
References

- Like It or Not: A Survey of Twitter Sentiment Analysis Methods
  (Authors: Anastasia Giachanou, Fabio Crestani)

- Sentiment Analysis and Opinion Mining
  (Author: Bing Liu)

- Google

- Twitter



Thank you