

A Survey on the Recent Techniques for Sentiment Analysis on Twitter

Beril Besbinar, Dimitrios Sarigiannis, Panayiotis Smeros (name.surname@epfl.ch)

École Polytechnique Fédérale de Lausanne (EPFL)

1. Problem

- Identify the **sentiment** of a group of people concerning a specific topic (based on hashtags and keywords)
- The input is a set of tweets and the output is their (positive or negative) sentiment

2. Use Cases

- **Prediction** of elections result
- Public opinion on global issues (e.g., refugees crisis, climate change)

Machine Learning

Algorithms

Promotion of products

3. Approach

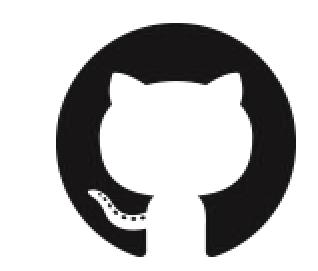
- Contractions Expansion (e.g., don't → do not)
- Emojis Transformation (e.g., :D \rightarrow <smile face>, :'(\rightarrow <sad face>)
- Emphasize Repeated Punctuation (e.g., yes!!!! → yes! <repeated>)
- Emphasize Repeated Last Characters (e.g., I am hapyyyyy → I am happy <repeated>)
- Filter Numerical Expressions (e.g., 35 → <number>)
- Split Hashtags (e.g., #lovemyjob → love my job) frequently used English words dictionary
- Emphasize Sentiment Words (e.g. I feel anxiety → I feel anxiety <negative>) SENTIMENT LEXICON

Text

Representaton

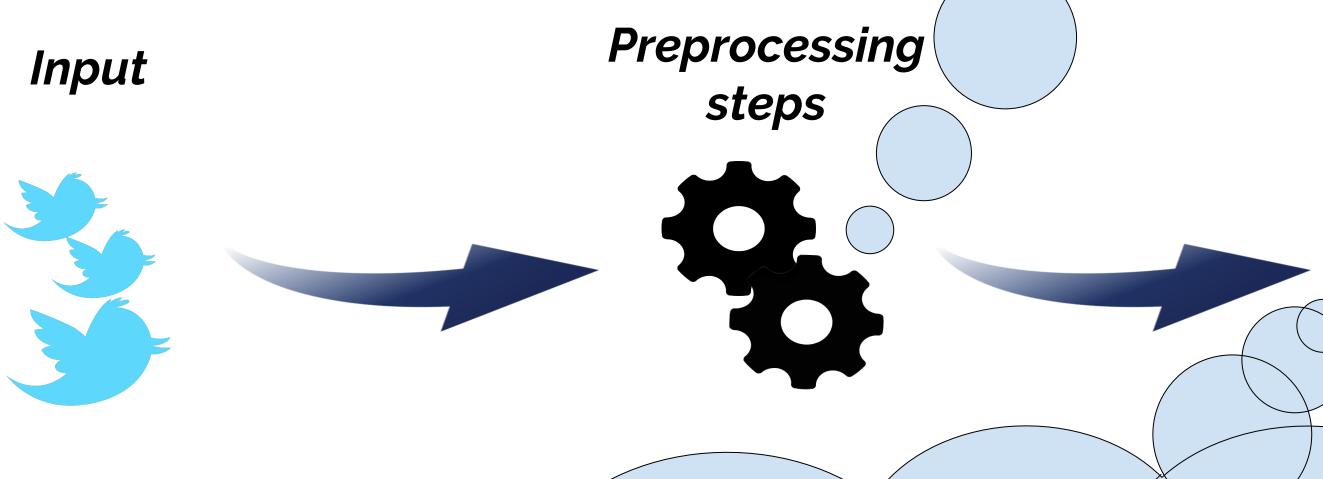
- Part-Of-Speech Tagging (e.g., guy → guy:Noun)
- Lemmatization and Stemming (e.g., followed → follow)
- Stop-words Filtering (e.g., the cat → cat)

Find Us





Sentiment



- pre-trained (GloVe)
- trained using GloVe algorithms
- hybrid

Tweet Embeddings

Word Embeddings

- concatenation
- summation
- averaging
- TF-IDF weighted

Bag of Words Representation

- boolean
- TF-IDF weighted
- using n-grams

Paragraph Embeddings

- distributed memory
- distributed bag of words

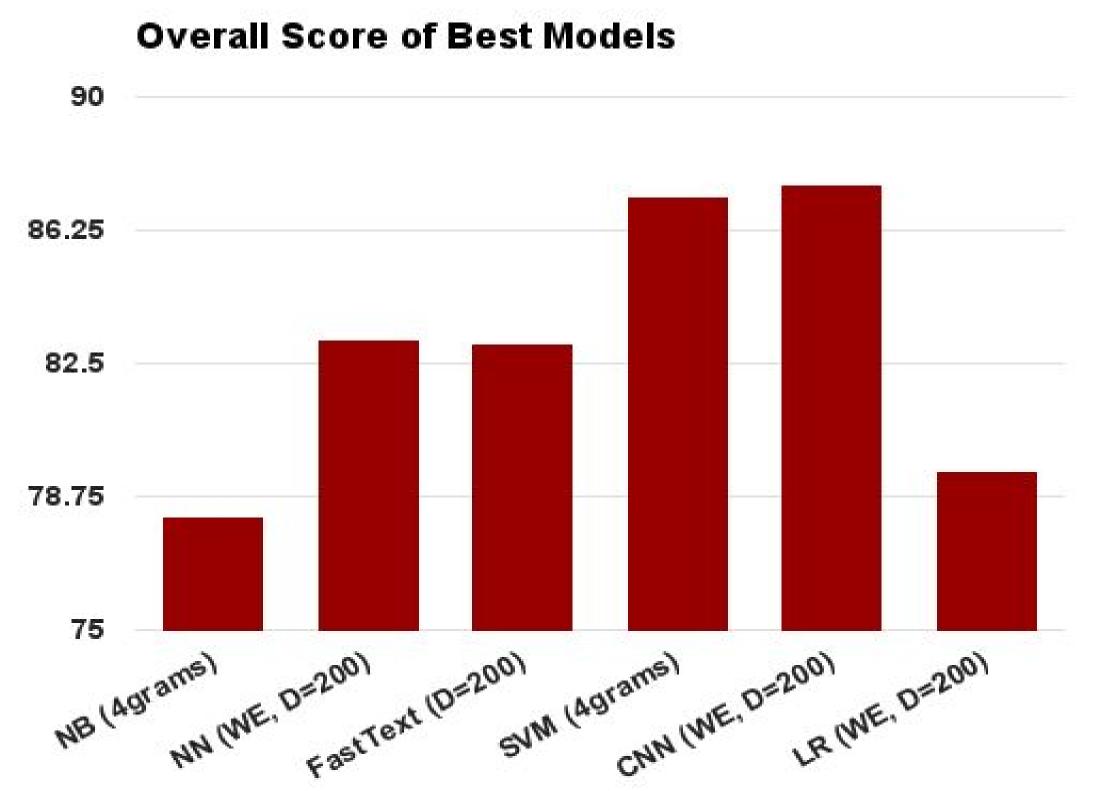
- Naive Bayes
- Logistic
- Regression
- FastText
- Support Vector
- Machines
- Convolutional

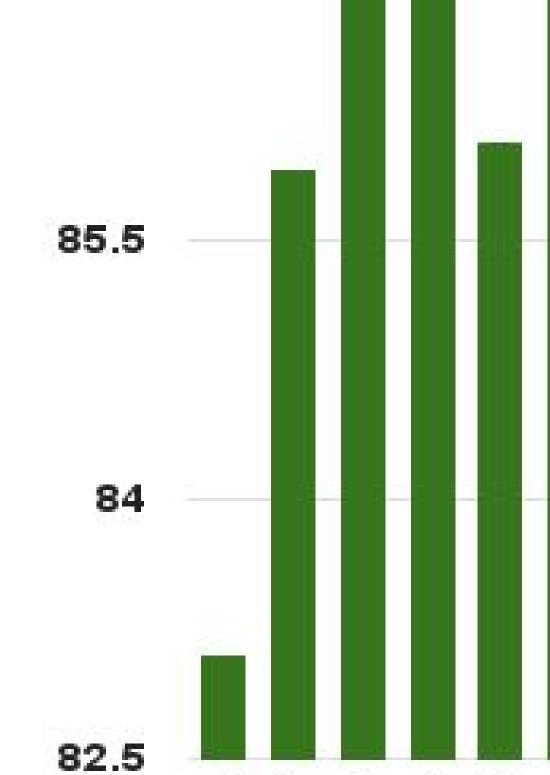
Neural Network



- Particle Swarm Model Selection (PSO)
- Convolution Max-pooling

N-grams Model Selection 4. Experiments WO 87





Fully-connected Gain without with Algorithm (from Preprocessing Preprocessing preprocessing) TF-IDF 85.28 87.20 1.92 (4grams) FastText 80.36 83.04 2.68 (D=200)

embeddings for each word