Toxic Comment Classification

MLoB

Outline

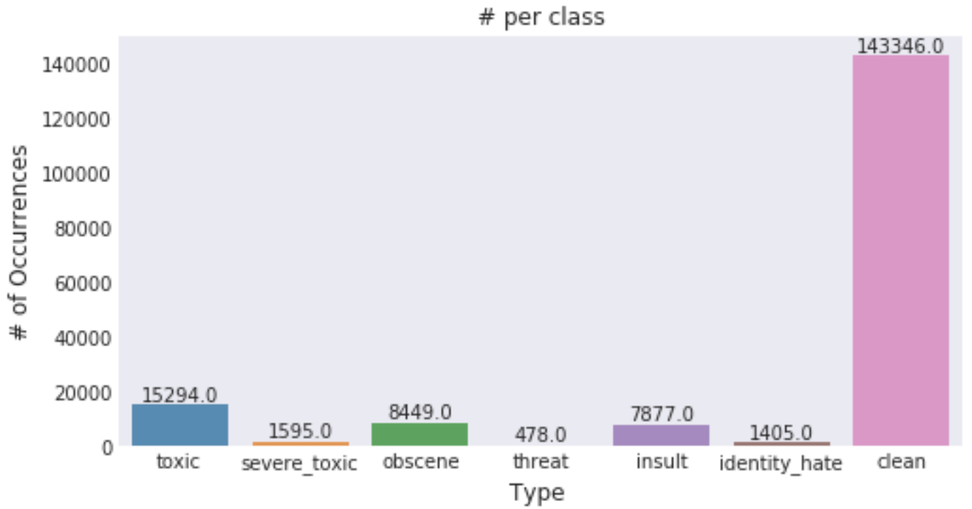
- Team introduction
- Competition
- Methods
- Results
- Discussion

Team

- Mick:
  - GCP
  - data cleaning
- Roel:
  - LSTM
  - Bi-directional LSTM
- Dennis:
  - Feature extraction,
  - GCP
- Joost:
  - Report,
  - Bi-directional LSTM
- Brian:
  - LSTM
  - Convolutional network

Toxic Comment Classification

- Multi label classification
- Imbalanced data set
  - Top x% vocabulary
  - Hand balancing
- ROC AUC
  - Label wise



Jagan, Stop the S@#$ - Toxic Comments EDA, https://www.kaggle.com/jagangupta/stop-the-s-toxic-comments-eda

Methods

- Feature models
  - SVM
  - MLP
- Independent models
  - LSTM
  - 1D convolutional network
  - Bi-directional LSTM

Feature Models

- Feature extraction
  - Ratio of capitals vs total characters
  - Ratio of punctuation characters
  - Total length in characters, words and in sentences
  - Total amount of some special characters: ?, (,), ! and some other characters.
  - Amount of unique words
- Expanding feature set
  - Swear words, word2vec tf-idf scores

- Dense Neural Network
- SVM
- Random Forest

Independent models

- LSTM
  - Input 902 LSTM Layer, Dense Layer
- Convolutional Network
  - Input 902 → 6 output, 2 convolutional layers (512 filters) , 2 Max Pooling Layers
- Bi-directional LSTM
  - Bi-directional LSTM layer, Max Pooling, 2x Dense Layer with drop-out

Loss -> Binary Cross Entropy

Optimizer -> Adam

Results

Method	AUC
Feature Based Approach v1	0.60
Feature Based Approach v2	0.53
Convolutional Neural Network	0.4902
Vanilla LSTM	0.52
Bidirectional LSTM	0.9256

- Kaggle final place : ***4090***

Looking forward

- Pre-processing
- Ensemble