



Predicting sarcasm on Reddit

Karl Meldorf, Mihkel Kruusi

Introduction

Reddit is a popular internet forum with dedicated subreddits for different subjects. Our main objective was to train a model, which can detect whether a comment is sarcastic or not. A problem, where humans are also frequently mistaken.

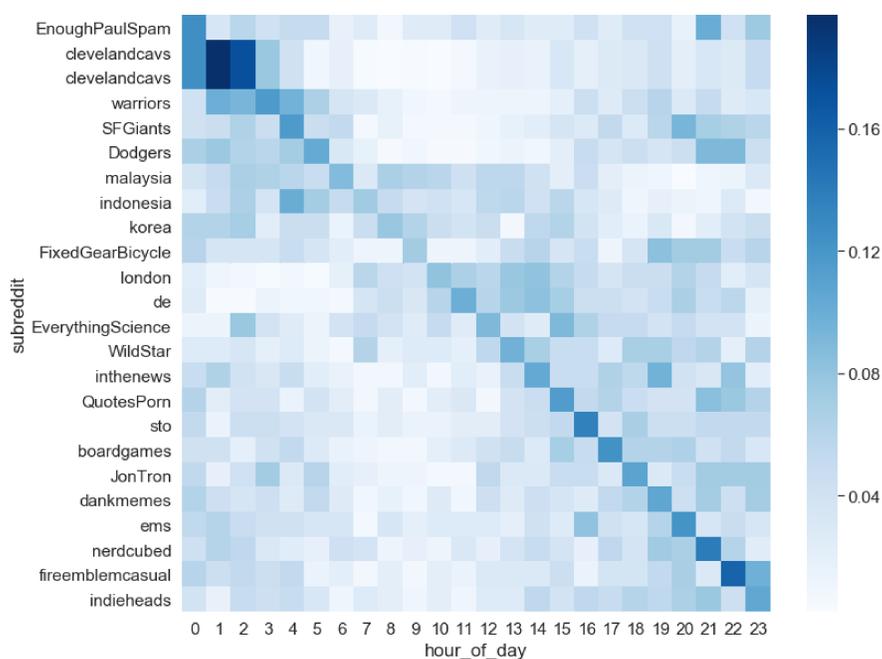
Importance/Motivation

Sarcasm detection is a part of natural language processing (NLP), which has recently gone from being only widely used to detect spam in email to smart assistants on smartphones and smart speakers, who understand natural language.

Data

We use a dataset published on Kaggle, which contains about 1 million rows and is balanced. Data is self annotated by users as on Reddit people can add a “/s” tag to their comment, which means the comment was sarcastic.

Cleaning the data we dropped comments, which contained a link, because sarcasm can be hidden in the link. Secondly we removed comments, which were shorter than 3 words on the assumption that people can't be sarcastic with less and if it is then is definitely harder to detect.



Top 3 subreddits in the data were AskReddit, politics and worldnews, but on the heatmap can be seen top subreddits by relative frequency by hour of day.

Training the model

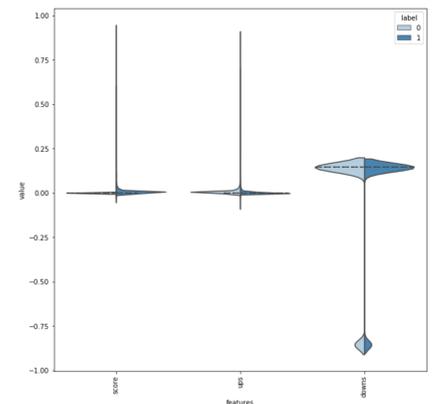
Classifier:

We used two common machine learning algorithms from Python Scikit-learn library: SVC with a linear kernel from support vector machines (SVM) and naïve Bayes classifier for multivariate Bernoulli models.

Comment features:

- *n*-grams (unigram, bigram)
- Sentiment

On the violin plot we have visualized label polarity of metadata, but from the symmetry we can deduce that metadata does not carry useful information for our classification model



Results

Results have to be taken with a grain of salt, because tests were ran on balanced data, but in reality people are sarcastic less frequently.

	precision	recall	f1-score
SVM* neg.	0.69	0.73	0.71
SVM* pos.	0.73	0.69	0.71
Naïve Bayes* neg.	0.70	0.64	0.67
Naïve Bayes* pos.	0.69	0.74	0.71

*features also include metadata (subreddit, score, ups, downs)

	precision	Recall	f1-score
SVM neg.	0.68	0.73	0.70
SVM pos.	0.73	0.68	0.70
Naïve Bayes neg.	0.68	0.64	0.66
Naïve Bayes pos.	0.69	0.72	0.70

SVM got slightly better results than Naïve Bayes and metadata slightly improved results, but only marginally meaning metadata didn't carry much information.

Conclusion

Sarcasm can be detected to some degree with simple machine learning and NLP tools as used in this project. For better results deep learning should be used with advanced methods to also understand context and meaning.