Twitter Virality Prediction

Benjamin Coors, Rémi Domingues, Håkan Eriksson, Paul Griffin, Akash Kumar Dhaka

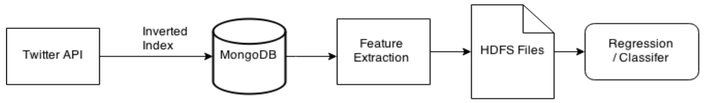
KTH Royal Institute Of Technology

1. Introduction

Twitter is a rich source of information. Mining it for data can be useful for a number of applications, for example in marketing. In our project we implemented a virality prediction of hashtags using machine learning and information retrieval techniques.

2. Methodology

2.1 Data pipeline



2.2 Features

- Followers
- Friends
- Listed
- Status
- Favourites
- Account age
- Verified
- Hashtags
- Media
- User mentions
- URLs
- Tweet length

2.3 Prediction Model

Two approaches:

1. **Regression:**

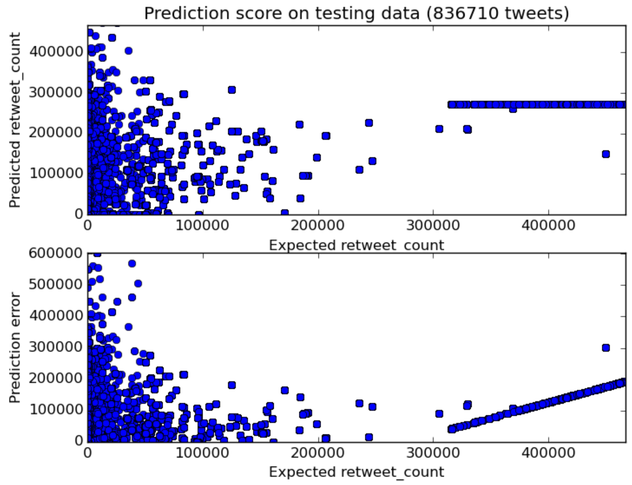
   Bayesian Ridge Regression

2. **Classification:**

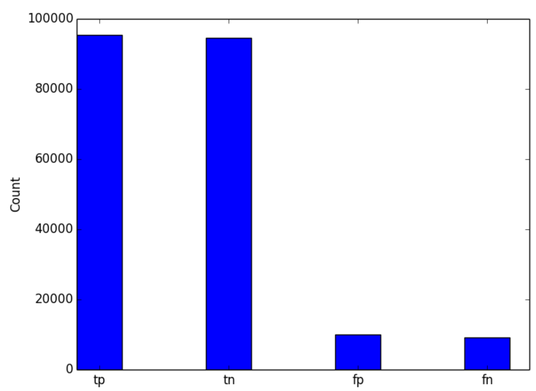
   Logistic Regression

3. Model performance

3.1 Regression



3.2 Classification



4. Results

4.1 Virality predictions

Top-10 predictions for the first May week:

Regression	Classification
BaltimoreRiots	BaltimoreRiots
BestFandomCA2015	BestFandomCA2015
FreddieGray	FreddieGray
Baltimore	Baltimore
NepalEarthquake	MGWV
SOL2015	NepalEarthquake
LIAMISTHEPOOLCL	FOLLOW
EANER	RT
HappyKyuziziDay	FOLLOWTRICK
HowIBecameAFan	Nepal
SaveMaryJaneVeloso	

5. Conclusion

The work done in this project shows the feasibility of predicting the virality of tweets using a feature-based approach. Future improvements could include further research into other fields such as social media analysis which will be needed to obtain even more accurate results.