A network tour of Data Science
Project proposal

Meryem Wehbe, Samuel Beuret & Valentine Santarelli

November 2016

Project description

Training a model

Our project consist in training a model analysing sentences based on the datasets described below. The model will also be evaluated according to this data (cross-validation, accuracy...).

Datasets:

1. The data set consist in sentences and their label i.e positive (score=1) or negative sentiment (score=0). `https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences` .
2. Amazon Product Reviews. In this dataset the comments of customers on products are labeled according to the rating they gave the same product. The rating is a good representation of the tone of the comment.
3. TripAdvisor reviews. Similar to the amazon product review dataset.
4. UMC Competition training data which contains polarity sentences.
5. More datasets might be used in case we need to improve our model

Test datasets and applications

The resulting model would be able to distinguish if a sentence would be negative or positive.
It would then be applied on different set of data.
We will parse facebook and twitter for posts of popular figures (using the corresponding API) and apply our model on the posts' comments to measure the reaction of the audience.