Readability and social networks

Ion Madrazo Azpiazu

Readability

- **Readability** refers to the ease with a reader can understand a given text.
  - The higher the score, the more complex and difficult a text it
- Supervised learning:
  - Example feature
    - Shallow features: Average words per sentence, average length of terms
    - Lexical features: Frequency of common/ uncommon vocabulary
    - Syntactic features: Syntactic tree depth, PoS frequencies
    - Semantic features: Concept density

Motivation and research plan

- Research has been done in readability for **standard/long texts**.
  - However, those methods tend to fail when assessing **short texts** like the ones in social networks
  - Could we create an specific readability tool for social media texts?

- Hypothesis
  - People tends to have contact with similar type of people
    - Similar readability people
  - Hypothesis: People pertaining to a same community will have a similar readability

Questions we aim to answer

- Improving readability
  - What does differ between communities, regarding the features we have for readability?
  - Are common readability features able to discriminate between two communities?
  - Can we obtain new features that do a better job at that discrimination?
  - Can those new features be used to improve general readability precision of our system? And if we specifically aim for short texts?
- Applications
  - Does a tweet's readability score have any impact in its probability of retweets?
    - Do people prefer to retweet tweets that have similar readability to their's
  - Does the diffusion of a tweet has any relation with its readability?
    - Do readable tweets get better diffusion that complex ones?