

Text Predictor Pitch

Jair G

The Text Predictor App

We gladly present the prototype of our text predictor application. Nowadays we spend a lot of time typing in our smartphones and whatnot and it is of the outmost importance to make this task easier and more efficient.

This application lets you predict the next word or the word you are currently typing offering the 5 most likely predictions using a nice presentation where the size of the word is proportional to the likelihood of the prediction. The aesthetics of the presented predicted words is based on the javascript library wordcloud.

Goals and Strategy of the Algorithm.

In building this application we focused on the following goals:

- ▶ Size on memory. The app has to use little resources.
- ▶ Execution Time. The app has to run quick enough to actually make predictions while you type.
- ▶ Performance. The predictions should be relevant.
- ▶ Presentation. The app should be visually appealing.

Data and Analysis.

The corpora used to train the algorithm comes from the *Heliohost Corpora* originally from <http://www.corpora.heliohost.org> and is composed of three corpus of text in US english: 1)Tweets from Twitter, 2)Blog entries and 3)News articles.

We used an N-gram model to build the prediction algorithm with the usual Markov Process assumptions. We constructed 4grams, 3grams, 2grams and 1grams of all possible combinations in the corpora and associated frequencies. We trimmed this tables to reduce the memory size needed without losing too much information; in fact, we reduced the size of the memory requirements by a factor of 10 keeping 60% of the cumulated frequencies in each corpus approximately.

Handling of Unseen Cases and Final Deployment

We used a simple backoff model to account for unseen cases nesting the 4gram, 3gram, 2gram and 1gram when the ngram was not observed in the corpora backing off to the next ngram. We did not use interpolation for this prototype.

The final app is deployed using the Rstudio's Shiny server and the App accomplishes satisfying results in relation to our intended goals.

You can see the prototype at

https://jairgs.shinyapps.io/text_predictor where you can find a complete documentation.