CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Conner J. Vercellino

Project Number

35353

Project Title

Jabberwocky: Use of Unsupervised Stacked Autoencoders for Next Generation Spell Checking Software for Dyslexics

Abstract

Objectives/Goals

Jabberwocky is a next generation spelling and grammar checking that uses stacked deep unsupervised autoencoders to make corrections from actual understanding of sentence context. This creates a far more accurate spell checker, comparable to a human editor. Using unsupervised text, it is able use a huge amount of data, removing the need for structured training data that most spell checkers need. It is able to learn individual or group spelling patterns to further augment the corrections, via structured datasets specific to the targeted user group.

Methods/Materials

The program needs to use entire sentences in order to be able to use the context of the sentence to help decipher what word the user intended to use. I looked into unsupervised models for neural networks. I decided on an noising filtering autoencoder based system. An noising filtering autoencoder is a type of neural network, whereby it tries to reproduce the input it is given without noise. Autoencoders have never been used for text correction making this project's approach to the problem of grammar and spelling correction very novel.

Results

Currently the network can make extremely accurate corrections on simple sentences. This proves that both an unsupervised neural network can work for spell correction, and that noise filtering does work on natural language. The program has been able to fix improper grammar and correct spelling on repeatedly mutated sentences. However, the network has yet to prove that it can generalize on new complex sentences that it hasn#t seen before.

Conclusions/Discussion

Jabberwocky has many uses, one such being that sites like Yelp, Airbnb and Wikipedia could be completely free of spelling and grammar errors, leading to the site#s content being more readable and professional, without effort of the site#s users. Search sites could use Jabberwocky on their queries, thereby removing the need to make their searches need to watch for improper use of spelling and grammar. Lastly, Jabberwocky, while currently trained for correcting text, could theoretically be adapted to facilitate code correction. This would be an extremely useful, as it could learn from existing code and it would be a huge boon to programmers.

Summary Statement

Using unsupervised stacked noise filtering autoencoders for creation of a extremely powerful NL correction algorithm (primarily aimed at Dyslexics).

Help Received

Classmate helped with the dataset loader; Andrew Saxe helped explaining about neural network training and math; Jen Selby helped with time management.

Preliminary