Abby Fisher

August 2, 2016

Professor Nicholas Flann

Independent Study

Documentation Report

News and Neural Networks (Jasper, 2015)

In order to have this code run correctly you must have node.js and python installed. If there is already data saved that you want to train with, simply type:

node process-data.js && python nn.py

Scrap-data.js:

This program scrapes Google News to build the data set that is used throughout the entire program.

Inputs: An array of queries used as search words and a series of dates that are used as a date range in the search.

The queries used:

queries = [ 'XOM' , 'MSFT', 'XOM', 'AAPL', 'JNJ', 'WFC', 'BRK', 'JPM', 'AT%26T', 'PFE', 'GE' ],

Outputs: Saves a data object that contains the full article separated into the body, headlines, date, and symbol the query used and puts each into their own object array.

doQuery(query)

Saves the data to the database for each query result

Run by typing: node scrap-data.js into the terminal or command prompt

Process-data.js

This program processes the data and creates the training.js file

Inputs:

Data set created from scrap-data.js

Outputs:

The training.js file

Run by typing node: process-data.js into the terminal or command prompt

nn.py

Uses training.js to create a training network and produces a prediction

Inputs:

Training.js vectorized data array

Outputs:

Accuracy of the training. Saves the trained data set into a folder within the program.

SupervisedDataSet( inp, target)

Data set that is used to train with. The inp used is the length of the training dataset.

BackpropTrainer(module, dataset=None, learningrate=0.01, lrdecay=1.0, momentum=0.0, verbose=False, batchlearning=False, weightdecay=0.0)

Trainer that trains the parameters of a module according to a supervised dataset (potentially sequential) by backpropagating the errors (through time).

Run by typing python nn.py into the terminal or command prompt.

Here are some results of the code by running it twice within 10 minutes of each other using exports4.js as the dataset.

The accuracy about doubled within the two runs of the program.

