



Week 8 Progress Report

Presenter - Matt Behnke

Mentor - Dr. Guo



Table of Contents

- Update on paper
 - Rough draft is currently complete, but references need to be added and cited in-text
 - I wanted to complete a whole attempt at the paper and then get feedback
- Neural Network results on which layers are being mis-identified
- Future Steps



Identifying which Titles are Incorrectly Labeled - Early Stopping Neural Network

- Slow algorithm, but worked
- Initially, create map of title to datafile
- Add all the mis-identified data files to a list after training and testing
- Then for every mis-identified file
 - Iterate through the map
 - For every datafile in the map, check every value in the data matrix of the value to see if it matches
 - If all individual data points match add , the title to a list of misidentified titles
- Enumerate titles for final result



Identifying which Titles are Incorrectly Labeled - Early Stopping Neural Network

- Over 10 trails of a 50/50 split for a neural network and added all instances of titles misidentified
- Layer 2 mis-identified more twice as much as the next highest (1: 10, 2:22)

```
Mis-Identified Titles by Layer
=====
Layer 2 mis-identified 22 times
Layer 1 mis-identified 10 times
Layer 14 mis-identified 9 times
Layer 44 mis-identified 8 times
Layer 16 mis-identified 7 times
Layer 6 mis-identified 6 times
Layer 3 mis-identified 6 times
Layer 28 mis-identified 5 times
Layer 4 mis-identified 4 times
Layer 5 mis-identified 3 times
Layer 59 mis-identified 2 times
Layer 43 mis-identified 2 times
Layer 8 mis-identified 2 times
Layer 7 mis-identified 2 times
Layer 58 mis-identified 1 times
Layer 53 mis-identified 1 times
Layer 49 mis-identified 1 times
Layer 48 mis-identified 1 times
Layer 46 mis-identified 1 times
Layer 45 mis-identified 1 times
Layer 42 mis-identified 1 times
Layer 34 mis-identified 1 times
Layer 33 mis-identified 1 times
Layer 17 mis-identified 1 times
Layer 15 mis-identified 1 times
Layer 10 mis-identified 1 times
=====
```



Future Steps

- Get Feedback and Revisions for Paper
- Interesting results for Early Stopping Neural Network mis-labels
 - More tests
 - Still need Random Forest to Work
 - Could include in results in paper if notable trends appear
- Work on final presentation for either Wednesday, Thursday, and Friday