

## **We run experiments on 5 words:**

### ***Serve***

Number of senses: 3

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1289303.001: 26

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1289303.014: 31

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1289303.005: 10

Total Count: 67

### ***Back***

Number of senses: 9

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1224262.020: 54

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1224262.018: 27

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1224262.014: 28

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1224262.023: 6

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1224262.001: 34

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1224262.008: 12

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1224262.035: 13

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1224262.022: 10

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1224262.038: 5

Total Count: 189

### ***Left***

Number of senses: 6

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1262701.001: 80

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1262701.004: 52

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1262701.010: 30

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1262777.001: 25

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1262777.006: 11

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1262777.012: 6

Total Count: 204

### ***Right***

Number of senses: 3

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1285123.022: 25

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1285123.014: 39

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1285123.002: 23

Total Count: 87

## **Open**

Number of senses: 8

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1273535.011: 23

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1273535.001: 37

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1273535.063: 2

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1273535.045: 10

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1273535.068: 11

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1273535.032: 6

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1273535.028: 9

Count for /dictionary/sense/en\_us\_NOAD3e\_2012/m\_en\_us1273535.024: 2

Total Count: 100

#####

For each word, we run a clustering algorithm and a neural network classifier.

## **Clustering**

We do NOT initialize cluster centers randomly. We take the average of each sense (average of sense's respective embedding instances) and let the K-Clustering algorithm begin optimizing from there. Thus, when calculating accuracy for clustering, we assume the clusters represent the gold standard sense classes. There is not train-test split here, we calculate clusters based on whole data and test on whole data as well.

## **Neural Network**

StratifiedShuffleSplit used for generating test-train split.

Train = 80%

Test = 20%

random\_state = 3

Structure:

Input Layer: (embedding dimension)

Hidden Layer: 30 units, 'relu' activation

Dropout Layer: 0.5

Output Layer: number of distinct senses, 'softmax' activation

SGD Optimizer, LearningRate = 0.005

Epochs = 80

BatchSize = 10

## Results

*Word - Clustering Accuracy - NeuralNet Accuracy*

### **Last 2 Layers Concatenated (dim=2048)**

Serve - 0.94 - 1.0

Back - 0.84 - 0.84

Left - 0.75 - 0.9

Right - 0.98 - 1.0

Open - 0.8 - 0.6

### **Last Layer (dim=1024)**

Serve - 0.85 - 1.0

Back - 0.8 - 0.79

Left - 0.79 - 0.9

Right - 0.97 - 1.0

Open - 0.85 - 0.55

### **Second-to-Last Layer (dim=1024)**

Serve - 0.94 - 1.0

Back - 0.84 - 0.86

Left - 0.77 - 0.87

Right - 0.98 - 1.0

Open - 0.84 - 0.75

## Notes

- Each run is done on the same test-train split to maintain consistency. Realistically, running a different/random train-test split each time will produce more variable results
- Senses with a count of 1 were removed
- The word 'right' contains more senses in the dataset, here it was only run on the 3 most frequent senses