

A BIG DATA APPROACH TO PREDICTING HELPFUL REVIEWS

Daniel Miller, Hector Castillo, Justin Stoner, Kelsie Box



CS 467/567 Principles and Applications of Big Data

Motivation

Sifting through review after review is time consuming yet necessary when it comes to online shopping. The motivation was to develop a method of sorting reviews based on how helpful they can be to the consumer. This way, consumers would only need to read a few reviews to feel confident in their purchase.

Clustering

- With clustering, we attempted to define what type of reviews Amazon buyers found helpful.
- We used K-means clustering with the silhouette method to determine how good our clusters were. Metric for “goodness” was how close an assigned point was to a neighboring cluster.
- Approach 1: Each row of data had a 35% chance of being included in a sample
- Approach 2: Rows with no helpful votes had

$$P = \tanh(\log_{10}(1) + 0.15) + \frac{r}{R}$$

of being included in a sample.
Rows with helpful votes had probability

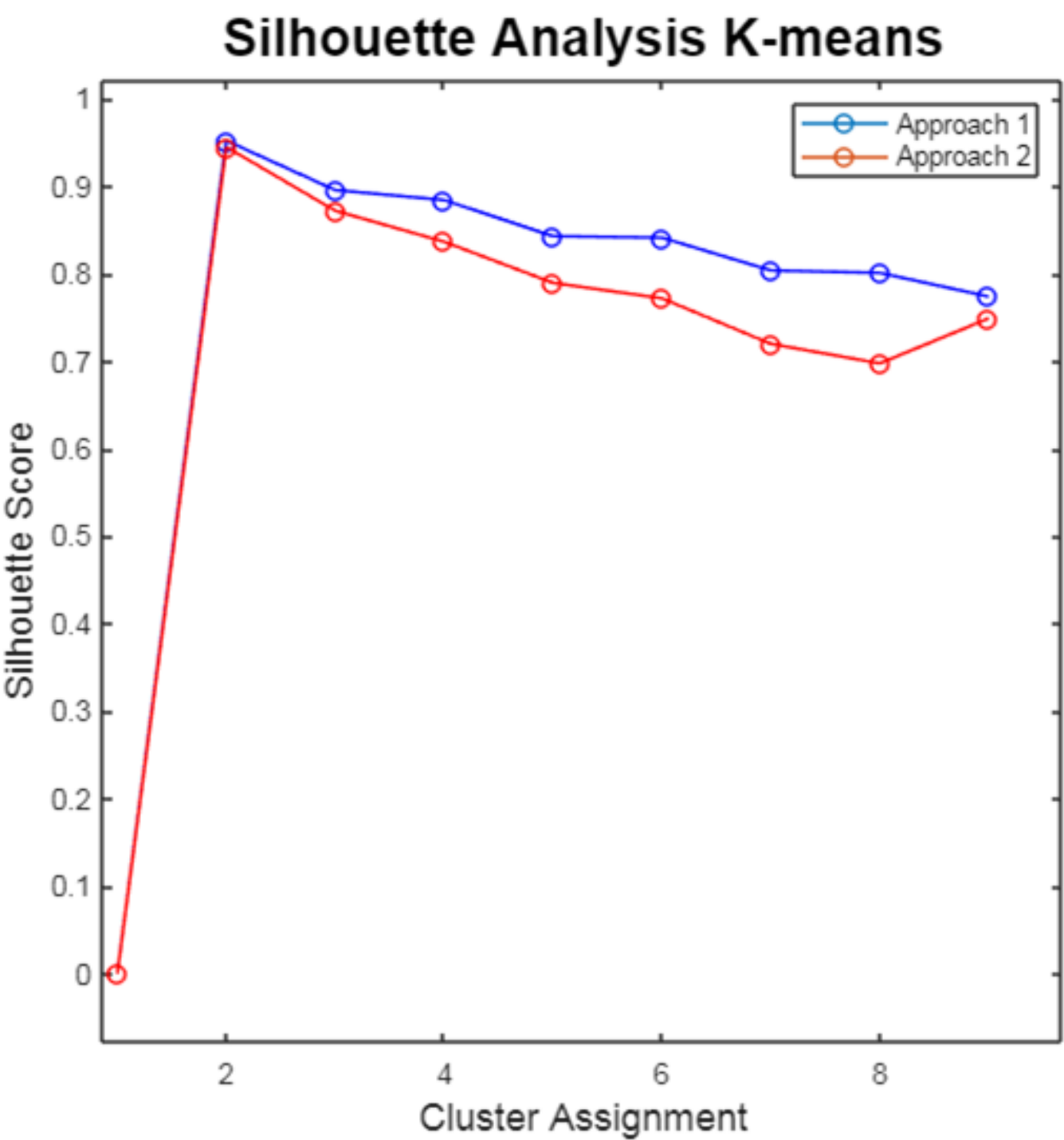
$$P = \tanh(\log_{10}(n) + 0.15) + \frac{r}{R}$$

of being included in the sample.

r : number of rows in the current data file.

R : total number of rows in all data files.

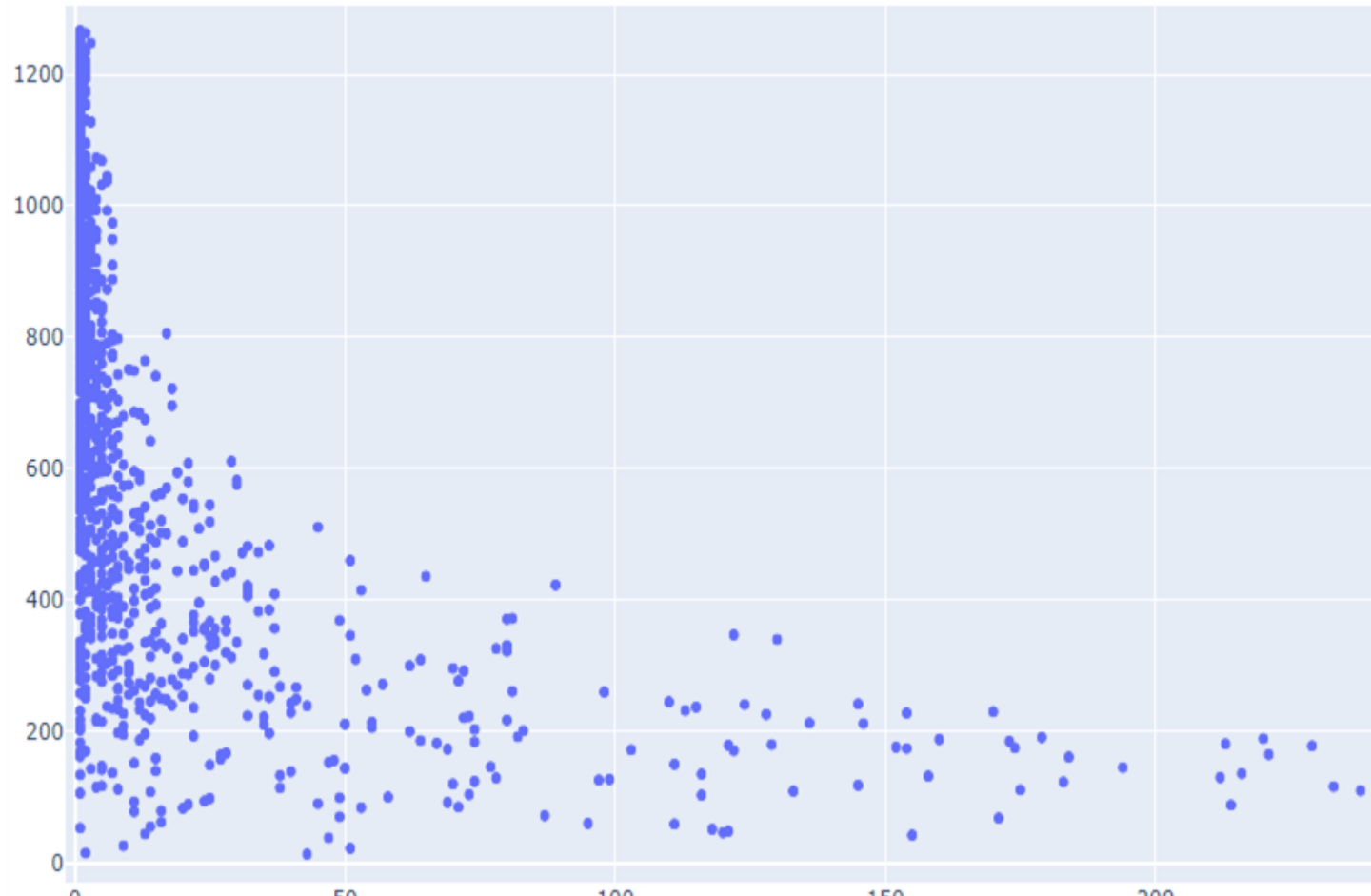
n : number of helpful votes for the review..



- Underlying goal is that the ideal number of clusters corresponds to the number of nodes used by the neural network for classification.
- Dimensions of the clustering space was Helpful Votes, Sentiment Rating, Subjectivity Rating, and Word Count.

Data

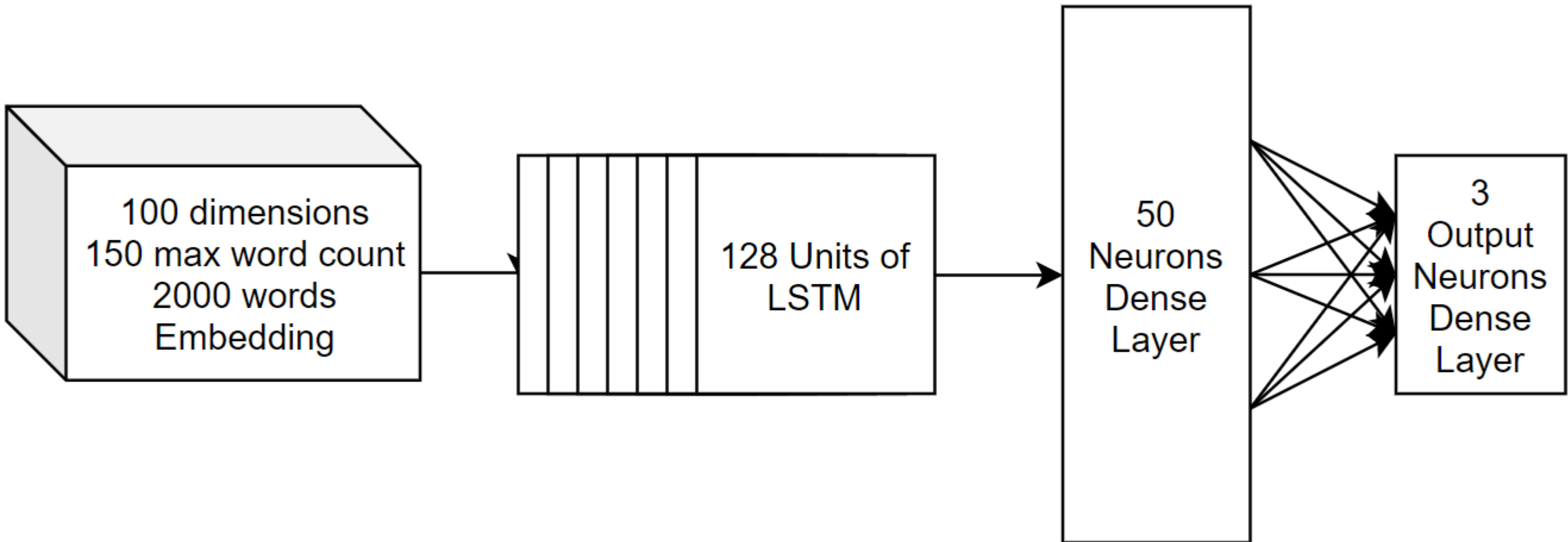
- It was compiled for over two decades since the first review in 1995.
- This dataset was constructed to represent a sample of customer evaluations and opinions, variation in the perception of a product across geographical regions, and promotional intent or bias in reviews.



(Helpful Votes, Number of Reviews)

Neural Networks

- Our first major design performed regression on the text to guess how many helpful votes a review should have. This had an extremely poor accuracy and forced us to changed our approach to find the helpfulness of a review.
- Our first working neural net model used convolution layers fed by embeddings and two dense layers of five units to guess what “tier” a review belonged in. Upon seeing that our neural net model classified everything as a tier 0 review, we again had to design the neural net.
- After much adjustment, we decided on 128 units of LSTM fed by embeddings and a dense layer of 50 neurons. We also reduced the number to tiers to three to get more control over the distribution of tiers. This design does not take the age of a review into account
- The final tiers are: less than or equal to 4 = tier 0, less than or equal to 16 = tier 1, everything else = tier 2



Results and Final Remarks

- In clustering method we found that the reviews that Amazon customers found the most helpful were those that were longer in length, neutral, and semi-objective. These tables represent the clusters with the highest average for helpful votes.

Table 1: Two Clusters					
	Helpful Votes	Sentiment Rating	Subjectivity Rating	Word Count	Sentence Count
avg	27.783	0.142	0.494	670.752	29.367

- As we increased the size of the clusters, we found that the number of helpful votes was inversely proportional to both the sentiment rating and subjectivity rating.

Table 2: Three Clusters					
	Helpful Votes	Sentiment Rating	Subjectivity Rating	Word Count	Sentence Count
avg	43.122	0.138	0.489	1121.853	47.228

- We did not find that the ideal number of clusters was the same as the number of nodes needed for classification in the NN.
- In the 2 cluster analysis, we found that reviews that were assigned in clusters with higher average helpful votes had on average 670 words, 30 sentences, 56 adjectives, 48 adverbs, 53 nouns, 55 pronouns, and 126 verbs.

Neural Net Accuracy

	Major Appliances	Gift Cards	Mobile Electronics
Validation Accuracy	0.7532	0.7740	0.6674

The final product of the neural net: The Configurations

For each file in the AWS data set, a file of weights and a tokenizer were produced. These can be loaded to evaluate the helpfulness of a review belonging to one of the product categories. The accuracy of these models generally falls between 70%-80%, the lowest being 66%.

For example, the tier 2 review “I read the negative reviews on this washer ... is bogus! never had a problem!” gives: [0.19202201 0.39777824 0.41019967], meaning it is most likely a tier 2 review

View our report so far and references

