TU Delft
Dina Chen
CSE3000
25-06-2020

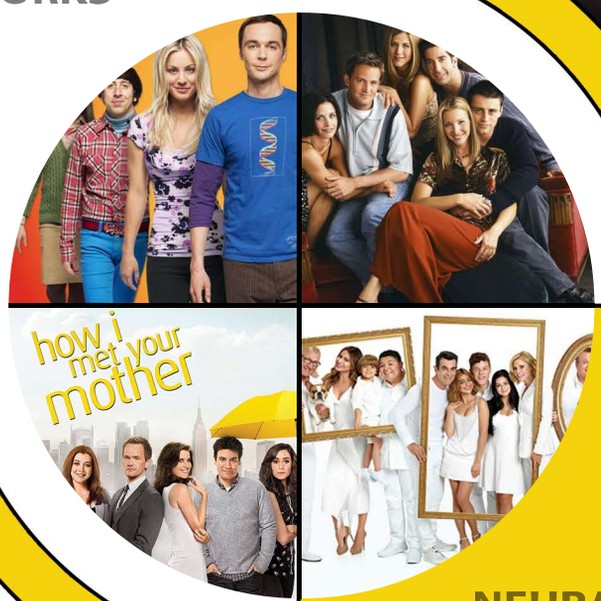
LOGISTIC REGRESSION

NEURAL NETWORKS

NEURAL NETWORKS

67%
SKIP-GRAM

58%



43%
SKIP-GRAM

47%
NEURAL NETWORKS

USING SKIP-GRAM MODEL

TO PREDICT FROM WHICH SHOW A GIVEN LINE IS

$$likelihood = \frac{1}{S}\sum_{s\in S}\sum_{j\in S, j\neq S} p(w_j \mid w_s)$$

$$p(w_O \mid w_1) = \frac{abs({v'_{wo}}^{\top} v_{w1})}{\sum_{w=1}^{W} abs({v'_{wo}}^{\top} v_{w1})}$$

SKIP GRAM MODEL TO PREDICT

METHODS

1. Generate Word2Vec embeddings for each show.
2. Calculate the likelihood (co-occurances of the words the sentence contains) for each show.
3. Balance them.
4. Take maximum as result.

LONG STORY SHORT

The resluts of cross-validation show that the accuracy of the best iteration of skip-gram is close to the best one of Logistic Regression Neural Networks.

The skip-gram model performs less stably, which leads to a lower average accuracy.

Removing stop words leads to better results.

SUPERVISORS:
Stavros Makrodimitris
Arman Naseri Jahfari
Tom Viering