Text generation for star wars character

Keshvi Srivastava
ksriva5@uic.edu

Varunya Yanamadala
vyanam3@uic.edu

Problem Statement

How to generate a sentence mimicking a star wars character with a specified mood?

1. Character/dramatis personae based text generation
   - Study various speech characters of each individual
   - Their preference for certain words or expressions
   - Eg. Yoda speaks in reverse english

2. Mood based text generation
   - Given a seed sentence, how well does the model pick up the mood and generate a continuous relevant sentence

Dataset

Star wars scripts:
Episode V
Episode VI
Episode VII

Kaggle
Star Wars scripts

IMSDB
Website containing all Star Wars movie scripts

Star wars HTML links::
Episode I
Episode II
Episode III

Pre-processing Data

1. Read '.txt' files and HTML links
2. Filter out characters
   a. Obiwan (Ben)   - Human :   583 dialogues
   b. Yoda   - Alien :   168 dialogues
   c. C3-PO (Threepio)   - Droid :   358 dialogues
3. Clean out the script
4. Tokenize the dialogues
   a. Generate a word vocab
   b. Assign a unique index to each word in the vocabulary list
   c. Convert the dialogues to their tokenized encoded versions
5. Create a 5-word padded encoded sequence to pass as input to the model

Model Outline

Seed sentence

Dialogue Sentence
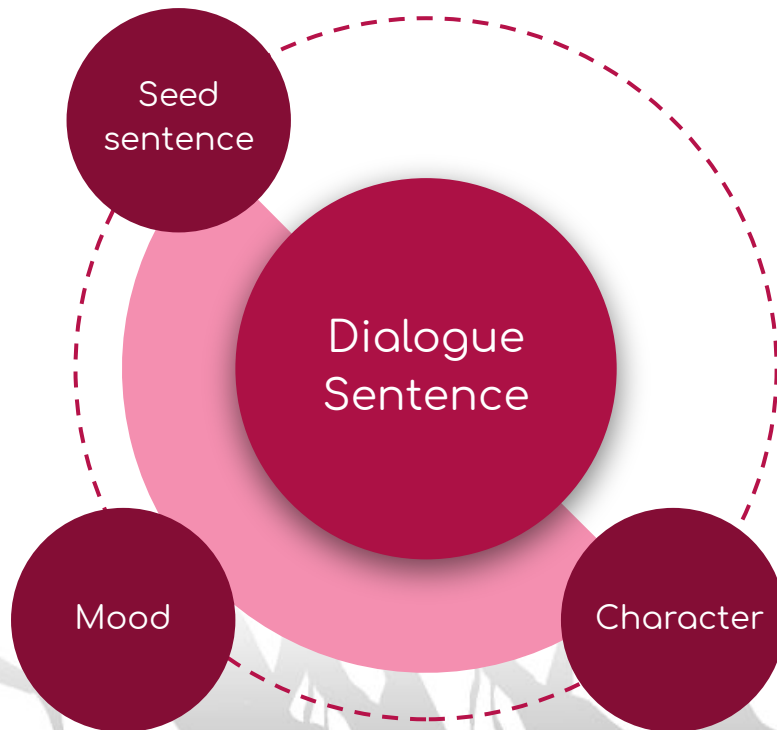
Mood

Character

→ Dialogue must be semantically correct

→ Avoid major grammatical errors

→ Should capture the uniqueness of the character to be mimicked

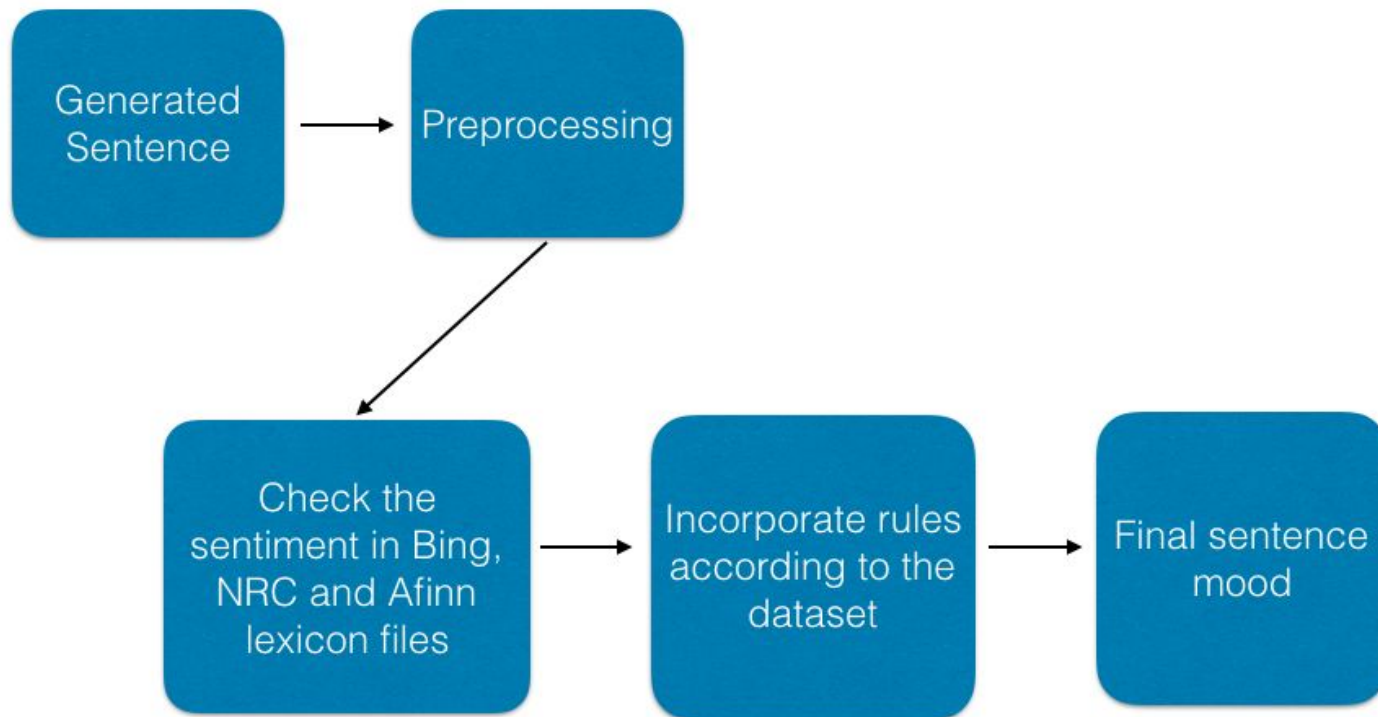# Model Outline – Text Generation

Bidirectional LSTM with GLOVE embeddings
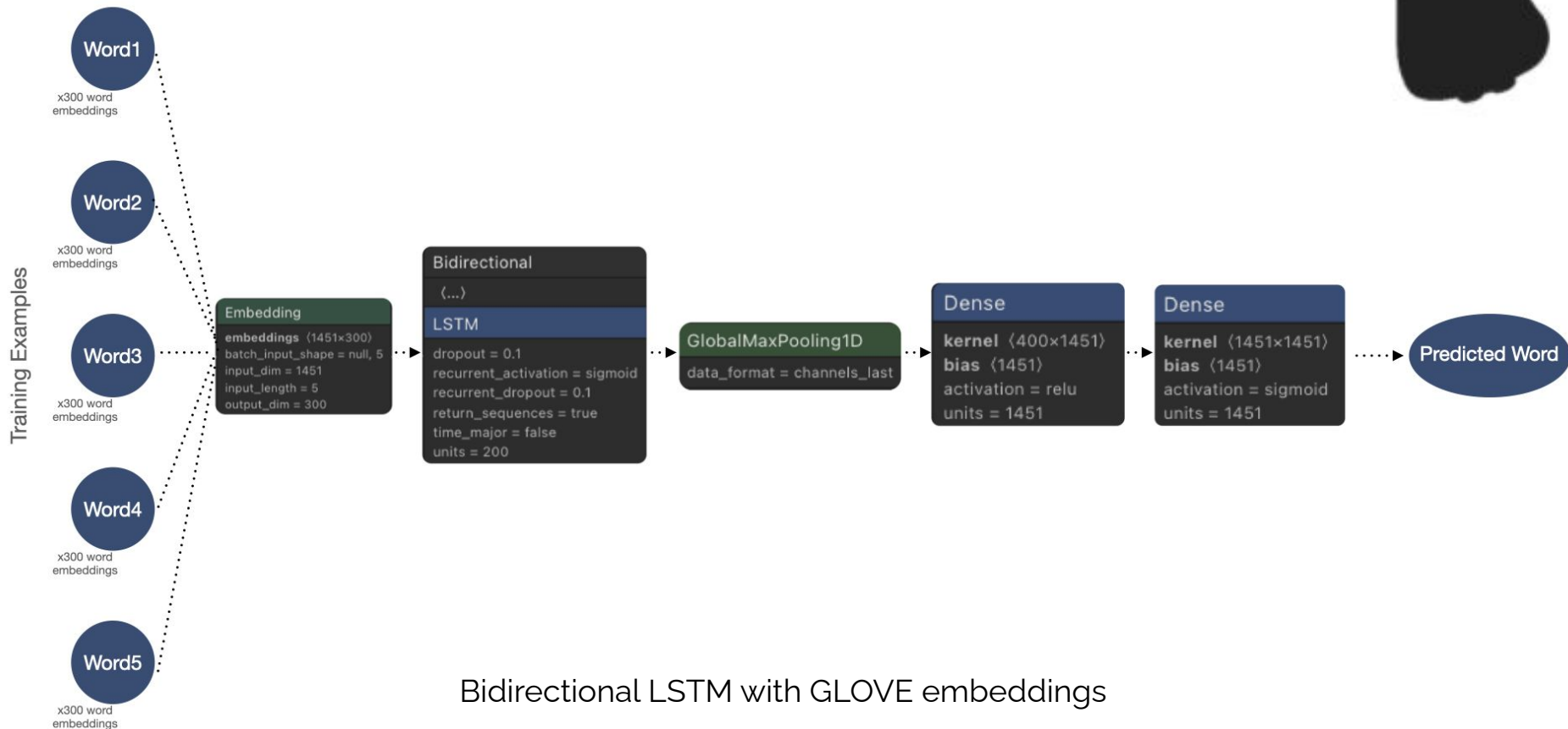
Generating Dialogues
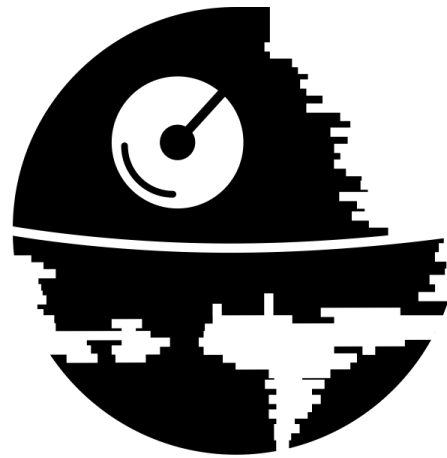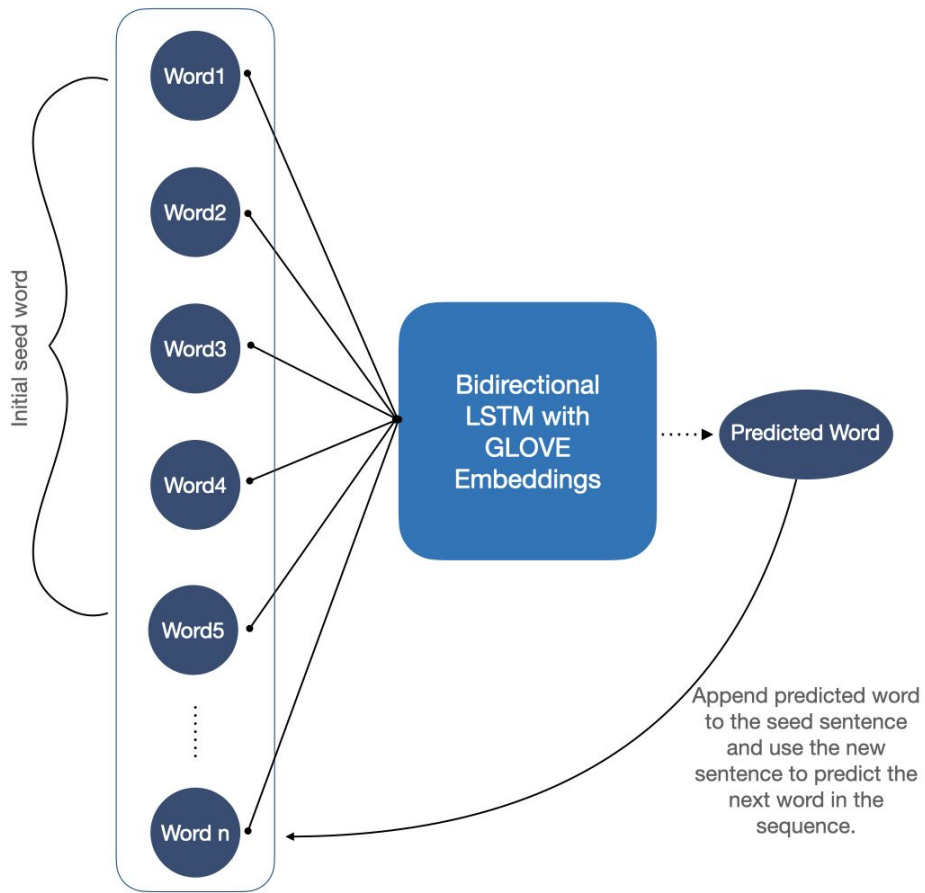
Model Outline – Mood Detection

Evaluation

1 We checked for the BLEU (Bilingual Evaluation Understudy) Score on the 10 generated dialogue for each character. The BLEU Score is calculated in reference to the training data used to train the model itself. It measure the similarity between the generated and original dialogues.

2 We also performed human evaluation on 5 out of the 10 generated dialogues for each character.
The survey was sectioned for each character. Each section included 5 dialogues with 3 questions for each.
Q1: How likely is the dialogue being spoken by that character?
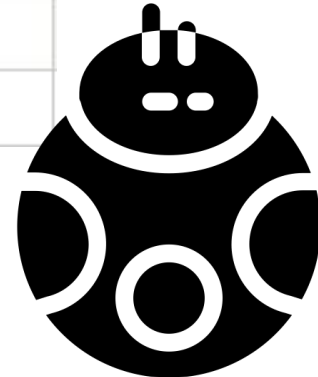Q2. Does the dialogue sound natural?
Q3. Does the dialogue make sense?

3 We checked for the mood of the seed dialogue and then the generated dialogue. We compare the two moods and see if the model is able pick the seed mood and generate the dialogue accordingly
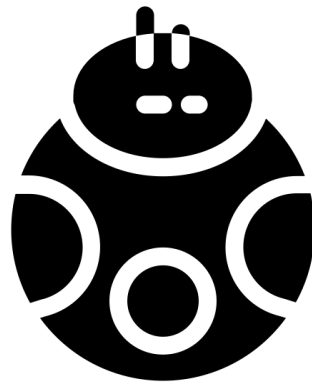
Results

|	Likeliness (How likely is the character to say the dialogue generated?) 1-5	Naturalness (How natural does generated dialogue sound?) 1-5	Sensibility (Is the sentence grammatically sensible?) 1-5	BLEU Score 0-1
Obi-wan	2.34	2.17	2.34	0.7051
Yoda	3.14	2.89	2.91	0.5636
C3PO	3.69	3.7	3.61	0.7685

Google Form Link :
https://docs.google.com/forms/d/1TnN4HxC6i1gJHcKvJDVfkTTbktq6GAcIyxjW00aaysA/edit?usp=sharing

Results

	Accuracy (Comparing the moods of generated sentences to the seed words)	From Survey
Obi-wan	55.70%	33%
Yoda	66.70%	33%
C3PO	77.80%	44%

Google Form Link:
https://forms.gle/6qFY6z8wCLbjqRvS6
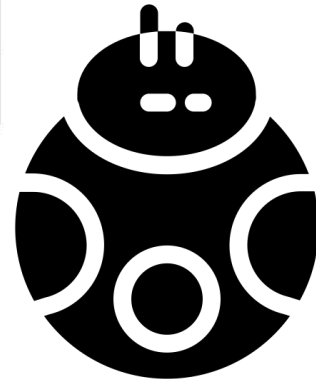
Learnings & Future Work

1. Lack of data:
   a. We observed that due to the lack of data, the model as a whole was not able to churn out the most meaningful sentences. While the sentences sounded like movie dialogues, they were not always grammatically sound.
   b. It would also be interesting to add more features to the model other than simple dialogues which could add as additional data and add more value to the characters dialogues.

2. Better model:
   a. GPT-2 and BART are two models which are newer to the field of text generation and would be interesting to dwell into for comparison.

Learnings & Future Work

3. Mood text generation:
    a. We proposed a model that would take a specific mood into account and generate sentence accordingly. However, our hypothesis was altered to assigning mood to a seed dialogue indirectly. We wanted to see if the model was capable of learning the mood without explicit seed.
    b. We can see from our results that this was not entirely successful.
    c. The classified mood for the seed was different from the generated one. This might be due to lack of keywords that help identify the specific mood. In this case the two moods predicted might be similar to one another or world apart.
    d. "A fully trained jedi knight the deepest force become you" would match with the seed words give as input whereas for "twilight is upon me and the dark side the dark" doesn't match with seed words mood.
    e. It would have been more appropriate to label each sentence of the dataset for training so as to check the relevance for the generated sentence mood classification. As an alternative, even a dataset which was labelled on a similar domain to leverage it over the generated sentences might be better.

Other Web References

- https://github.com/monologg/GoEmotions-pytorch
- https://imsdb.com/search.php
- https://www.kaggle.com/xvivancos/star-wars-movie-scripts?select=SW_EpisodeIV.txt
- https://homes.cs.washington.edu/~msap/movie-bias/
- https://aaai.org/ocs/index.php/AIIDE/AIIDE17/paper/viewFile/15832/15250
- https://medium.com/sciforce/a-comprehensive-guide-to-natural-language-generation-dd63a4b6e548
- https://medium.com/@b.terryjack/natural-language-generation-nlg-fe8844db6f01
- https://medium.com/swlh/star-wars-an-nlp-adventure-5e0deb70cd75
- https://medium.com/deep-writing/how-to-write-with-artificial-intelligence-45747ed073c#.n9zpfkh60
- https://medium.com/@plusepsilon/the-bidirectional-language-model-1f3961d1fb27
- https://www.tensorflow.org/tutorials/text/text_generation#the_prediction_task
- https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/text
- https://towardsdatascience.com/generating-short-star-wars-text-with-lstms-c7dc65e252c4
- https://towardsdatascience.com/film-script-generation-with-gpt-2-58601b00d371
- https://towardsdatascience.com/how-to-evaluate-text-generation-models-metrics-for-automatic-evaluation-of-nlp-models-e1c251b04ec1
- https://towardsdatascience.com/simple-text-generation-d1c93f43f340
- https://arxiv.org/pdf/2006.14799.pdf

Other Web References

- https://machinelearningmastery.com/calculate-bleu-score-for-text-python/
- https://machinelearningmastery.com/how-to-develop-a-word-level-neural-language-model-in-keras/
- https://machinelearningmastery.com/text-generation-lstm-recurrent-neural-networks-python-keras/
- https://github.com/aherriot/story-generator
- https://github.com/FernandoLpz/Text-Generation-BiLSTM-PyTorch
- https://github.com/amisha-jodhani/text-generator-harry-potter
- https://www.kaggle.com/hamishdickson/weighted-word-autocomplete-using-star-wars-dataset
- https://www.kaggle.com/guidant/mimicking-star-wars-characters-using-a-i-rnn#2.-Data-Preparation
- https://www.kaggle.com/liopic/classic-nlp-exploration-of-star-wars-dialogues
- https://www.kaggle.com/andrebarros1/starwars-chatbot-nlp-application/output
- https://www.kaggle.com/xvivancos/analyzing-star-wars-movie-scripts/execution
- https://www.kaggle.com/eriktrs/exerc-cio-chatbot-star-wars
- https://www.kaggle.com/shivamb/beginners-guide-to-text-generation-using-lstms
- https://stackabuse.com/python-for-nlp-deep-learning-text-generation-with-keras/
- https://www.analyticsvidhya.com/blog/2018/03/text-generation-using-python-nlp/

May the Force be with you.