

Summer Odyssey SPMS 2021 – Smart Email Assistant

Presented by Teo Yong Sen

Supervised by Prof Cheong Siew Ann

Introduction

Motivation:

- Advancements in technology and communication tools enable the ease of usage for emails
- Corporations use emails as the main means of communication

Method:

- Natural Language Processing (NLP) tools like POS Tags and Chunking.
- Create a Deep learning model that classify emails into 4 different email classes.
- Used Categorical Cross-entropy Loss function for model.

$$J(\mathbf{w}) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

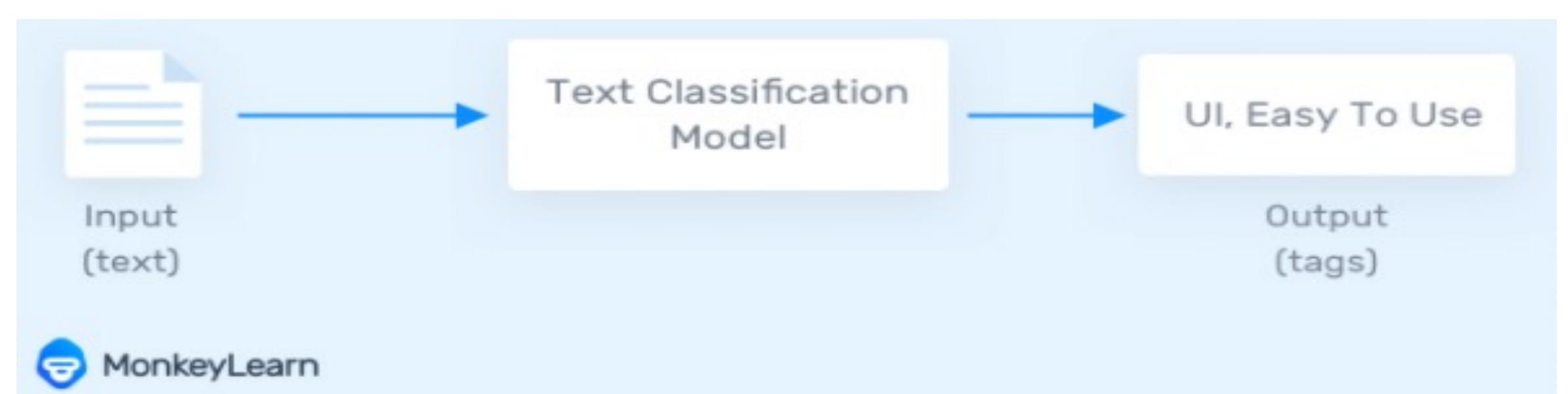
Results:

- We achieved a Categorical Cross entropy loss of 0.65 and an accuracy of 75%.

Features

- We used text vectorization.
- Each email is converted into a fixed sequence length consisting of integers, with each integer representing each word in the email.
- Integers are indices of words in the global word vocabulary

Model Design

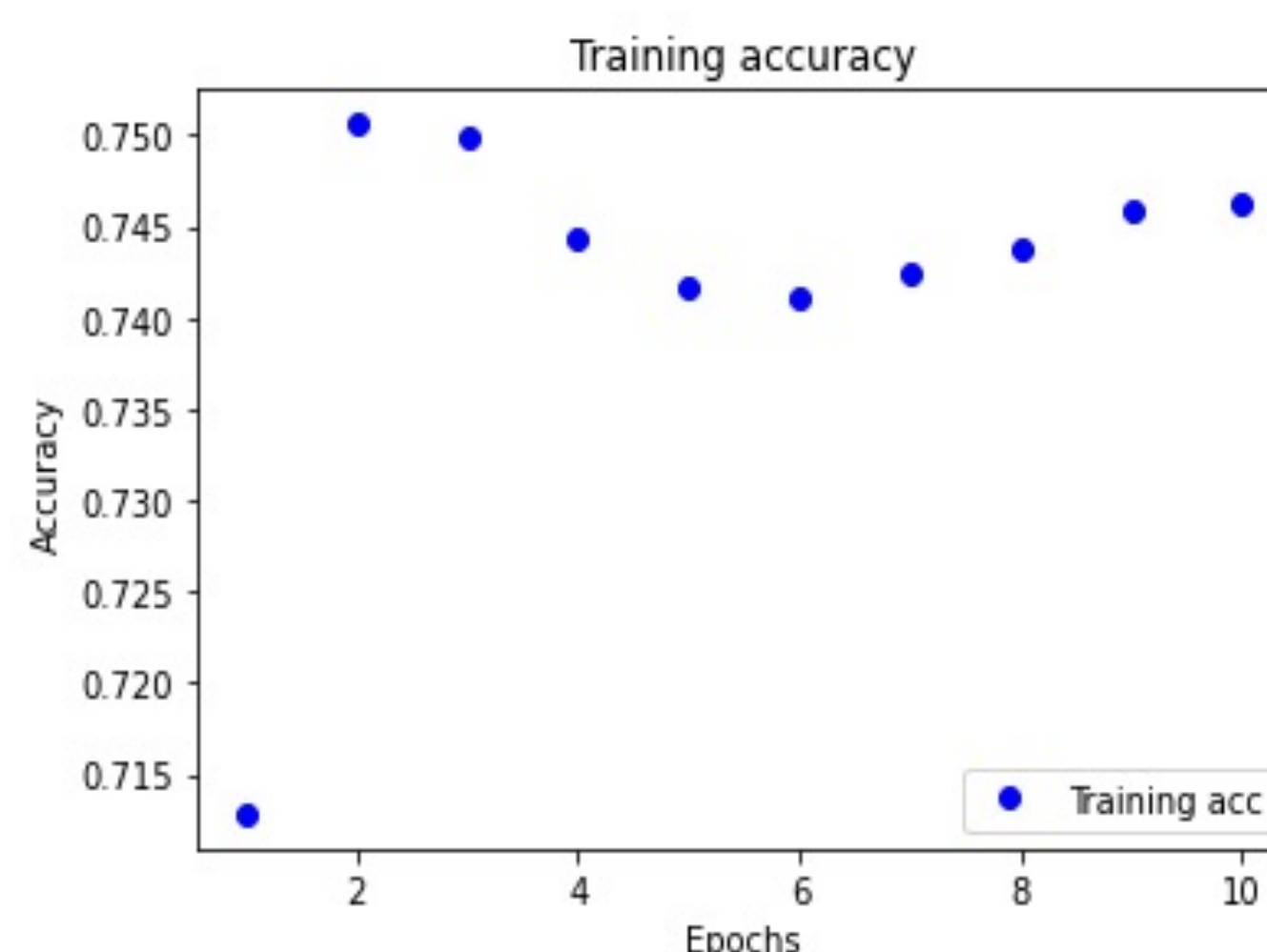


Data

- Model was trained using only email content and not the subject of the emails.
- Email Classes that we created:
 - (1) – Need reply, need action (31938 emails)
 - (2) – Need action, no reply (111578 emails)
 - (3) – Need reply, no action (9572 emails)
 - (4) – Information only (43915 emails)

- Created a vocabulary bank consisting of all unique words from emails dataset.
- Each word is tokenized, and each email is vectorized to feed into Neural Network.

Results



10 epochs ran with highest accuracy of 75% shown at 3rd epoch

Discussion

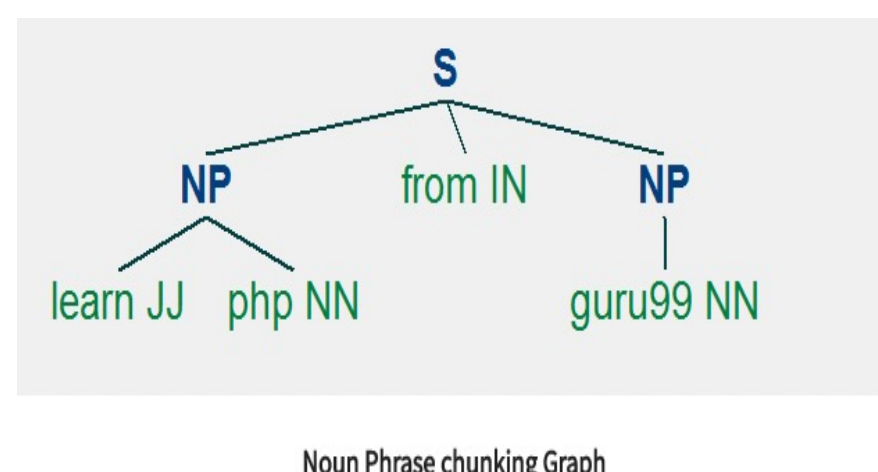
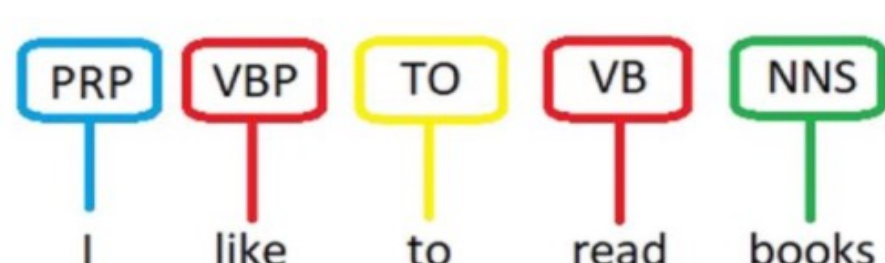
- Future improvements can be done by removing stopwords and regularization of the weights in the model.
- Unsupervised learning can also be done using TensorFlow and compare the results with supervised learning.

Exploratory Stage

Parts of speech tags

Chunking

POS Tagging



- POS tagging involves tagging each word to a particular tag.
- Examples of tags are: Personal Pronoun, Verb, Noun
- Chunking involves extracting phrases out of sentences
- Examples of chunks are Noun Phrase, Verb Phrase
- Each sentence is made from phrase chunks, which in turn consists of POS tags.

References:

- Mehta, D. (2021 Jan). Part Of Speech Tagging – POS Tagging in NLP. Byteiota. <https://byteiota.com/pos-tagging/>
- Monkeylearn (2021). Text Classification. Monkeylearn. <https://monkeylearn.com/text-classification/>
- Guru99. POS Tagging with NLTK and Chunking in NLP [Examples]. Guru99. <https://www.guru99.com/pos-tagging-chunking-nltk.html>