

Deep Neural Network for Sentence Classification

Devanshi Shah
dept. ICET
Centennial College
Toronto, Canada

dshah163@my.centennialcollege.ca

Hitesh Dharmadhikari
dept. ICET
Centennial College
Toronto, Canada

hdharma1@my.centennialcollege.ca

Jefil Tasna John Mohan
dept. ICET
Centennial College
Toronto, Canada

jjohanmo@my.centennialcollege.ca

Shrikant Kale
dept. ICET
Centennial College
Toronto, Canada

skale5@my.centennialcollege.ca

ABSTRACT

Sentence classification is extremely useful for several use cases such as document classification, spam filtering, and sentiment analysis. Existing models for sentence classification that use neural networks do not take into account the context in which they appear. However, even without the context, these models are useful for tasks such as providing semantic headings to unstructured text. In this paper, we describe the experiments performed to develop an architecture for sentence classification using natural language processing and deep sequence modeling. Our model reaches an optimal performance by leveraging a state-of-the-art model for generating text embeddings.

I. INTRODUCTION

A technology that can assist a user in quickly locating the information of interest is highly desired, as it may reduce the time required to locate relevant information. Researchers often skim through abstracts in order to quickly determine if the paper suits their interest. This process is easier when abstracts are structured, i.e., the text in an abstract is divided into headings such as the objective, method, result, and conclusion. However, a significant portion of abstracts or information is unstructured, which makes it more difficult to quickly access the information of interest. Therefore, classifying each sentence of an abstract into an appropriate heading can significantly reduce the time to locate the desired information. So, in order to convert unstructured data to a

structure formative, we propose a solution which is a sentence classification task, in order to distinguish it from general text classification or sentence classification that does not have any context. Our project uses Deep Neural Network Architecture for sentence classification.

II. DATASET

We are evaluating our model on the sentence classification task using the PubMed medical abstract datasets, where each sentence of the abstract is annotated with one label. In total, we are splitting the sentence into major 5 labels. The dataset consists of 200k out of which we are considering 20k values. The dataset has three major files: test.txt, train.txt, and dev.txt. PubMed 20k RCT assembled this corpus consisting of randomized controlled trials (RCTs) from the PubMed database of the biomedical literature, which provides a standard set of 5 sentence labels: objectives, background, methods, results, and conclusions.

III. METHODOLOGY

Three neural network models were developed and evaluated to determine the optimal architecture for the sentence classification task. A baseline model using Naïve Bayes was also developed to evaluate the performance against all future models.

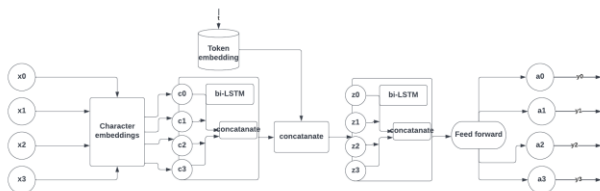
1. SUPERVISED LEARNING MODEL

Text Vectorization and text embedding is performed on the data before it is used for training the model. Text vectorizer converts the text into numbers and text embedding is performed so that model learns the relationship between tokens in our data. One hot encoded label are provided as the target for the model. The model architecture consists of 1D convolutional layer with 'GlobalAveragePooling' and 'relu' as the activation function. A 'softmax' activation function is used in the final dense layer and the model is Compiled with Adam optimizer and a loss function of 'categorical_crossentropy'.

2. UNSUPERVISED LEARNING MODEL

The second experiment involved using pre-trained token embeddings from TensorFlow Hub for feature extraction. A universal sentence encoder from TensorFlow Hub is utilized for the task of generating token embeddings. The architecture of the model includes a feature extractor model using the TF Hub layer which generates the embeddings. The output of this layer is used as input to the next Dense layer with 128 hidden units and the 'relu' activation function. A 'softmax' activation function is used in the final dense layer and the model is compiled using an 'Adam' optimizer and 'categorical_crossentropy' as the loss function.

3. STATE-OF-THE-ART-MODEL



The final experiment of our model consists of three components: a hybrid token embedding layer, a positional embedding layer, and an optimization layer. The first component consists of combining the pre-trained token embeddings and the character level embedding to form a hybrid embedding component. The second component involved using positional token embedding and concatenating with the previous hybrid embeddings to form a tribid embedding. The final component of the model is a feed-forward layer which takes as input the sequence of vectors from the previous layer and outputs a sequence of labels. The score of a label sequence is defined as the sum of probabilities of individual labels and transitional probabilities. The scores are then turned into probabilities of the label sequence by taking an activation function 'softmax' over the all-possible label sequences.

IV. RESULT

Table 1 compares the performance of 4 models developed during the experimentation. The models are evaluated on classification performance metrics such as precision, recall, f1 score, and accuracy.

Metrics	Baseline model	Model1	Model2	Model3
Precision	0.78	0.79	0.73	0.83
Recall	0.72	0.78	0.71	0.83
F1-score	0.74	0.79	0.72	0.83
Accuracy	0.72	0.78	0.72	0.83

Table 1: Model performance for sentence classification on the validation set

Table 2 shows the performance of the state-of-the-art model on the testing set for various classification metrics.

Metrics	Score
Precision	0.83
Recall	0.83
F1-score	0.82
Accuracy	0.83

Table 2: State-of-the-art model performance on the testing set

V. CONTRIBUTION

Task	Contributors
Data pre-processing	Devanshi Shah, Jefil Tasna John Mohan, Hitesh Dharmadhikari, Shrikant kale
Base Multinomial Model	Hitesh Dharmadhikari
Supervised Model	Jefil Tasna John Mohan
Unsupervised Model	Shrikant Kale
State-of-art Model	Devanshi Shah
Documentation (Report and Presentation)	Devanshi Shah, Jefil Tasna John Mohan, Hitesh Dharmadhikari, Shrikant Kale
Research and Analysis	Devanshi Shah, Jefil Tasna John Mohan, Hitesh Dharmadhikari, Shrikant Kale

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].
- [5] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.