RMDL: Random Multi Model Deep Learning for text Classification

Garima Malik

April 21, 2021

OUTLINE

- Introduction
- Literature Review
- RMDL Architecture
- Project Overview
- Methodology
- Results

Ryerson University

INTRODUCTION

- Project Paper : "RMDL: Random Multimodel Deep Learning for Classification" released in 2018.
- Github Repo : RMDL Source Code.
- Python Package : "! pip install RMDL"
- NLP_Project Repo : My Project Code

Ryerson University

LITERATURE REVIEW

RMDL paper organized the Literature Review in 3 parts :

- **Feature Extraction**
  - L. Krueger et. al.[1] proposed feature extraction methods based on word counting for text categorization in statistical learning
- **Classification Methods and Techniques**
  - K. Murphy [2],I. Rish [3] discussed about Naive Bayes Classifier and its empirical analysis
  - C. Yu et.al [4],S. Tong et. al.[5] proposed SVM algorithm with active learning and latent variable techniques.
- **Deep Learning methods for Text Classification**
  - D. Cires [6] proposed multi column deep neural networks for classification tasks especially visual recognition tasks.
  - K. Kowsari et. al.[7] implemented HDLTex: Hierarchical Deep learning for text classification

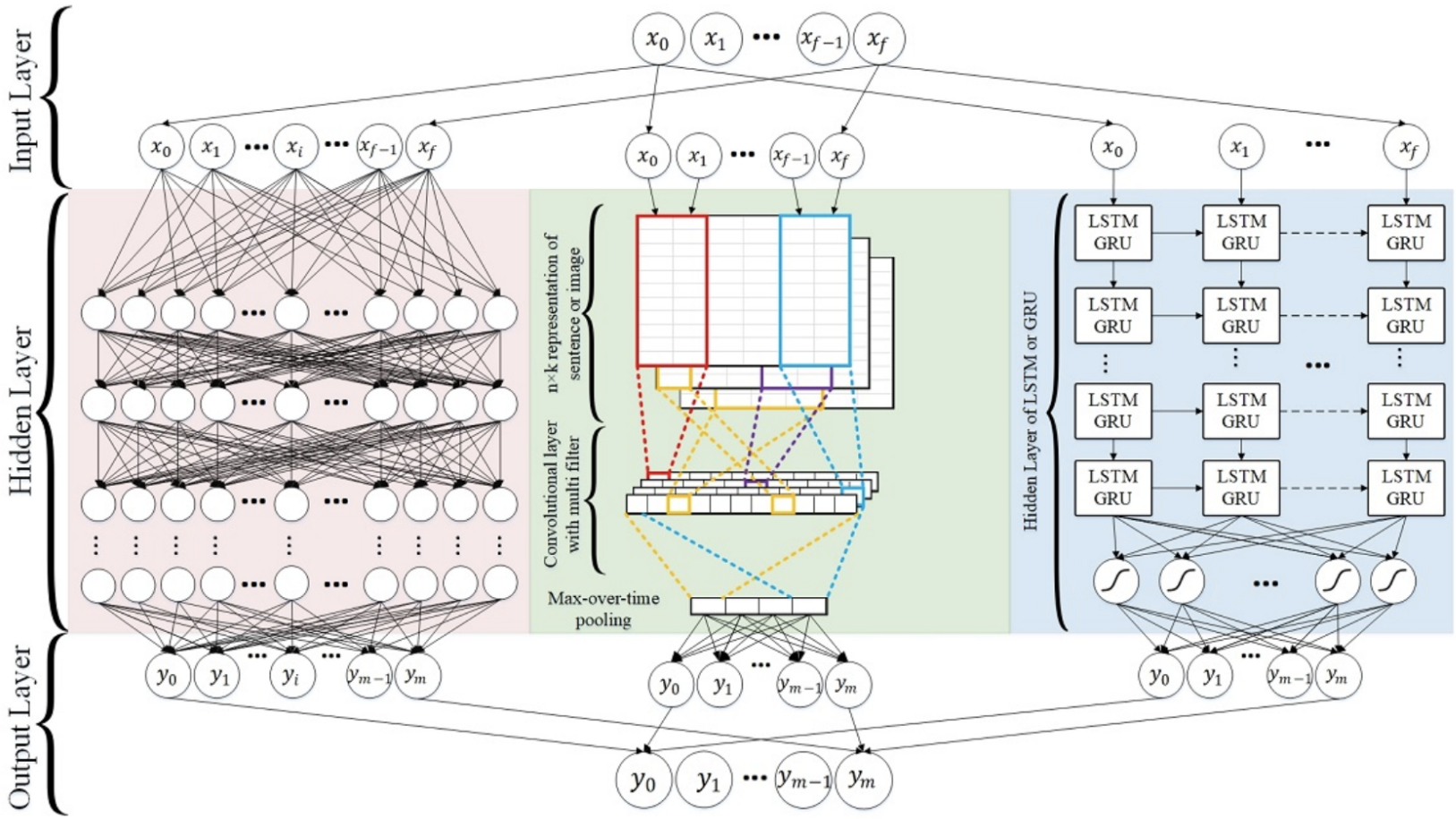
Ryerson University

RMDL: Random Multi Model Deep learning for Classification

- **Research Motivation :**
  - The continually increasing number of complex data sets each year necessitates ever improving machine learning methods for robust and accurate categorization of these data.
  - Users need to manually do hyper parameter tuning by changing each and every parameter which results into longer execution times.
- **Proposed Solution in Paper :**
  - The proposed approach uses a basic concept of randomization.
  - It tries to randomize each parameter of deep learning models and gives the best possible combination of parameters with each DNN, RNN and CNN models

Ryerson University

RMDL Architecture



Project Overview

- **Aim of the Project:**
  - To empirically analyse the feasibility of the proposed Model in the paper "RMDL: Random Multi model Deep Learning for Classification" in comparison with Roberta BERT Model.
- **My Contributions:**
  - Replicated the RMDL model architecture on IMDB and Reuters Data set
  - To Assess the effectiveness of RMDL model on recent text classification data set, created a text classification data set (scrapped from Kaggle) called as Stack Overflow question/answer data set.
  - Trained Roberta BERT Model using simple transformers library.
  - Compared the performance of RMDL and Roberta BERT model over all the above-mentioned data sets.

Ryerson University

PROJECT OVERVIEW

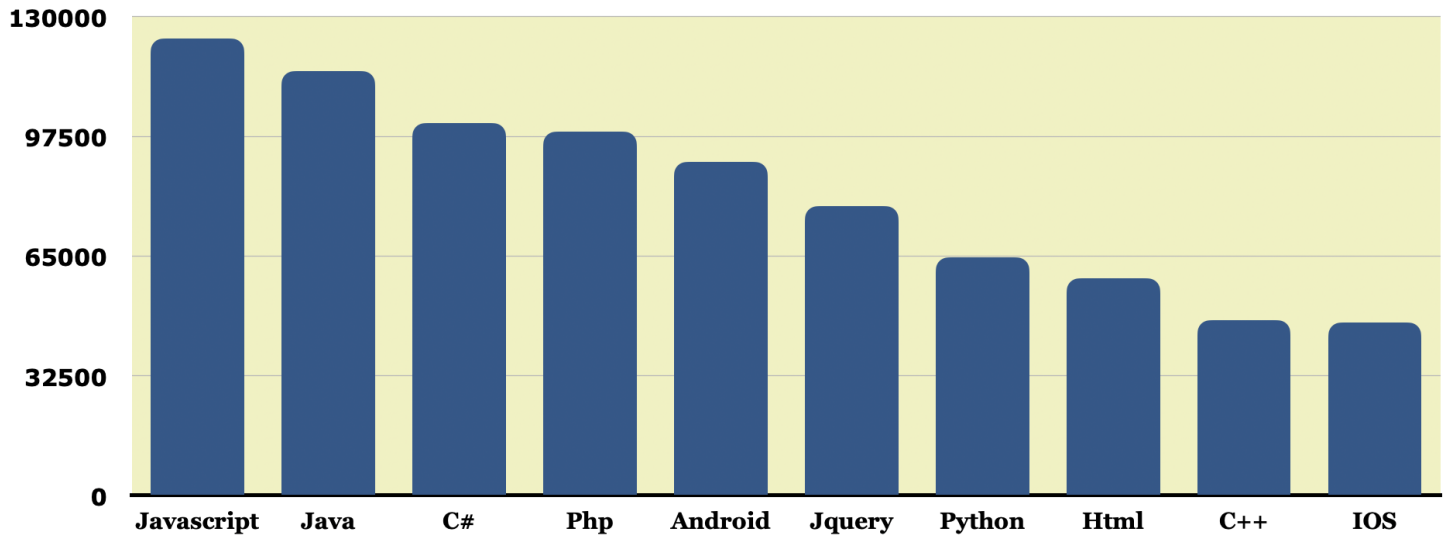


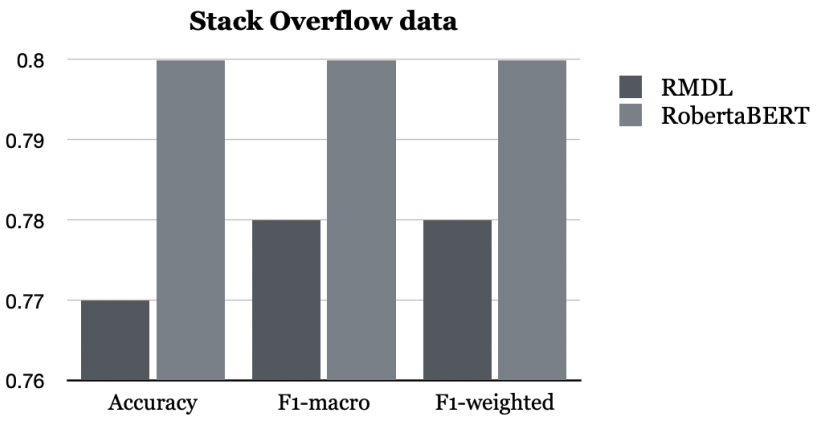
Figure: Showing the Label distribution in stack overflow data set

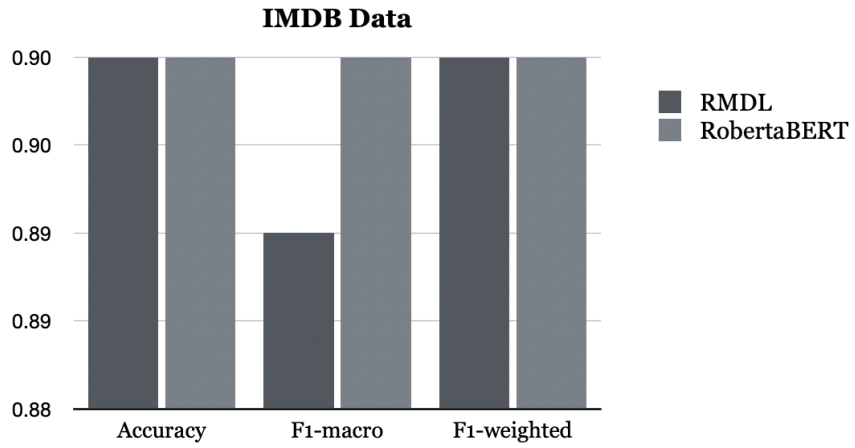
```
from simpletransformers.classification import ClassificationModel

#Create a ClassificationModel
model = ClassificationModel('roberta', 'roberta-base', num_labels=2, use_cuda = False)
```

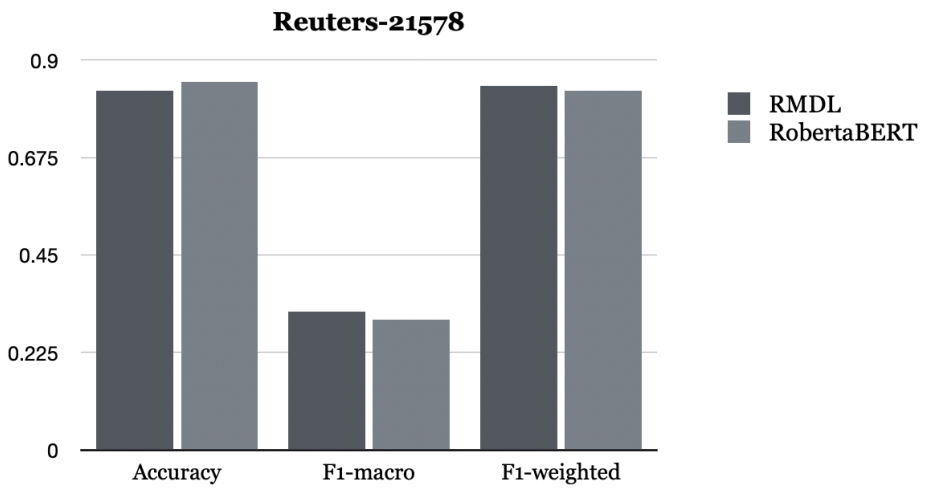
Figure: Code Snippet to train the Roberta BERT model

Ryerson University

RESULTS


Stack Overflow data
0.8
0.79
0.78
0.77
0.76
Accuracy
F1-macro
F1-weighted
RMDL
RobertaBERT


IMDB Data
0.90
0.90
0.89
0.89
0.88
Accuracy
F1-macro
F1-weighted
RMDL
RobertaBERT


Reuters-21578
0.9
0.675
0.45
0.225
0
Accuracy
F1-macro
F1-weighted
RMDL
RobertaBERT

CONCLUSION

- **Limitations of RMDL:**
  - Longer Execution Times
  - Excessive Randomization
- **Results Discussion :**
  - In case of stack overflow dataset : BERT model outperformes RMDL models with an f1-score of 80%
  - In case of IMDB dataset : Both models perform equally well
  - In case of Reuters dataset : RMDL perform slightly better than BERT
- **Future Work :**
  - Extensive experimentation is required for better assessment as the execution time is high for both the models for a data set of more than 10,000 rows.
  - RMDL can be implemented with new set of embeddings and feature sets such as ELMo, BERT and fastText

Ryerson University

REFERENCES

1 Krueger, L. E., Shapiro, R. G. (1979). Letter detection with rapid serial visual presentation: Evidence against word superiority at feature extraction. Journal of Experimental Psychology: Human Perception and Performance, 5(4), 657.

2 Murphy, K. P. (2006). Naive bayes classifiers. University of British Columbia, 18(60)

3 Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence

4 Yu, C. N. J., Joachims, T. (2009, June). Learning structural svms with latent variables. In Proceedings of the 26th annual international conference on machine learning (pp. 1169-1176).

5 ong, S., Koller, D. (2001). Support vector machine active learning with applications to text classification. Journal of machine learning research, 2(Nov), 45-66.

6 CireAan, D., Meier, U., Masci, J., Schmidhuber, J. (2012). Multi-column deep neural network for traffic sign classification. Neural networks, 32, 333-338.

7 Kowsari, K., Brown,Hdltex: Hierarchical deep learning for text classification. In 2017 16th IEEE international conference on machine learning and applications (ICMLA) (pp. 364-371). IEEE

Ryerson University