Multi Label Text Classification Using Transformers

Alekya Paladugu (302845283)

Samah Eltayeb (302740659)

Srujay Reddy (303207866)

Introduction

Classification tasks in machine learning can be categorised into Binary Classification, Multiclass Classification, and Multi-label Classification.

Binary and Multiclass Classifications deal with mutually exclusive class assignments, Multi-label Classification presents a more complex scenario where an instance may belong to multiple classes simultaneously.

Our focus is on enhancing the efficiency of Multi-label Classification models. We chose the Toxic Comment dataset for its challenging nature, encompassing 159,571 samples across six nuanced labels.

Problem Formulation

Multi-Label Text Classification (MLTC) poses a unique challenge in accurately assigning multiple labels to a single input sample. This complexity is compounded by the interdependencies among labels, a factor often overlooked in conventional models. Our project proposes a model that effectively captures these label dependencies, aiming to surpass the performance of existing models.

To approach this task, we implemented various models, starting with classical machine learning models, followed by deep learning models like BiLSTM and CNN, and then progressing to more advanced models like MAGNET (Multi-Label Text Classification using Graph Neural Network with Attention). Our contributions is the standalone BERT transformer model.

System Architecture:

Our project incorporates a range of classical and deep learning algorithms.

Models Implemented:

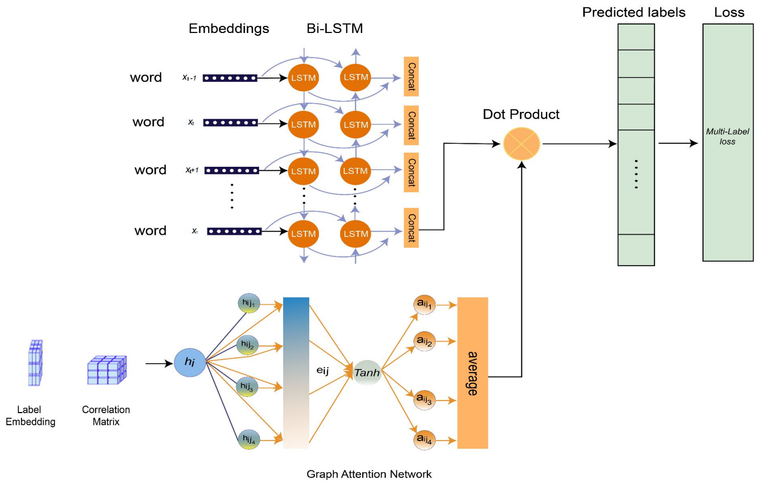
- Binary Relevance.
- Classifier Chain.
- One Vs Rest
- Label PowerSet.
- Hierarchical SVM.
- BiLSTM using glove and BERT embeddings.
- CNN using glove and BERT embeddings.
- MAGNET (Attention-based Graph Neural Network).
- BERT model.

Methodology

- Preprocessed the Toxic Comment Dataset by removing stopwords, making the dataset less biased.
- Splitted the data into 80% for training and 20% for testing using train_test_split method.
- Implemented different models ranging from classical machine learning models, MAGNET and BERT.
- Our aim was to benchmark the BERT model against the baseline MAGNET model.

Baseline Model: MAGNET

- It is an amalgamation of two distinct models. In this setup, the Bidirectional Long Short-Term Memory (BiLSTM) network is utilised for generating feature vectors, taking word embeddings as input.
- Concurrently, the Graph Attention Network processes the Adjacency matrix and label vectors, outputting label-specific features.
- These label features are then integrated with the feature vectors produced by the BiLSTM, creating a comprehensive model that effectively captures the nuances of multi-label text classification.



Best Model: BERT

- We leveraged the pre-trained BERT model, adapting it to our multi-label classification context through specialised preprocessing, classification layers, and evaluation methods.
- Used tokenization to break the input text into individual words.
- Fine-tuned the BERT model on our dataset by adding a classification head on top of the core BERT model.

Results

The BERT model has the highest accuracy of 0.68 and beats the baseline MAGNET model.

	model_name	micro_avg_f1_score	hamming_loss	accuracy
0	BR	0.758944	0.077863	0.661787
1	Classifier_Chain	0.764628	0.077581	0.668259
2	One_vs_rest	0.759005	0.077838	0.661787
3	Label_Powerset	0.722120	0.085362	0.653621
4	Hierarchical_SVM	0.775363	0.073523	0.671957
5	BiLSTM_glove	0.824693	0.077262	0.658860
6	BiLSTM_bert	0.827579	0.076183	0.659630
7	CNN_glove	0.815251	0.080101	0.652234
8	CNN_bert	0.814277	0.079947	0.660092
9	MAGNET_Cooccurance	0.820034	0.078935	0.640216
10	MAGNET_xavier	0.818782	0.079177	0.644838
11	MAGNET_random	0.815133	0.081532	0.630200
12	BERT	0.856885	0.065247	0.684615

Thank You!!