In this research study. a standardized length of 256 has been implemented. The parameter "maxlen" has been configured to 256 to signify the extent of comment sentences. This adjustment facilitates the truncation of sentences exceeding 256 words in the text comments while preserving their length at 256. For sentences with fewer than 256 words. zero-padding is applied to extend them to the specified length. Zero padding is incorporated at the outset until the comment length reaches the maximum. maintaining uniformity. Sentence length padding involves the introduction of zero tensors (placeholders devoid of information) to prevent the occurrence of empty elements and ensure consistency. The addition of a zero tensor at the beginning achieves the same outcome as appending it at the end.

Dropout is a technique that helps prevent overfitting and allows for efficient combination of numerous neural network topologies [1]. It is evident that the optimal performance is achieved when the dropout rate is set to 0.3. This is because a dropout rate of 0.3 results in the most random generation of network structures. We experimented several numbers of learning rates. i.e., 1e-3, 5e-5, and 2e-5. Whereas for the batch size, we chose three numbers: 16, 32, 64, and 128. Table 1 shows the results of all experiments using BiLSTM+AugGPT that have been carried out.

Table 1. The performance comparison results for BiLSTM+AugGPT (identification)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Batch Size** | | | |
| **Learning rate** | **16** | **32** | **64** | **128** |
| Code Comment (CC) | | | | |
| 1e-3 | **0.939** | 0.933 | 0.935 | 0.937 |
| 5e-5 | 0.936 | 0.936 | 0.928 | 0.935 |
| 2e-5 | 0.936 | 0.936 | 0.934 | 0.931 |
| Issue Section (IS) | | | | |
| 1e-3 | **0.878** | 0.878 | 0.878 | 0.874 |
| 5e-5 | 0.844 | 0.876 | 0.870 | 0.856 |
| 2e-5 | 0.868 | 0.870 | 0.856 | 0.855 |
| Pull Section (PS) | | | | |
| 1e-3 | **0.862** | 0.853 | 0.832 | 0.821 |
| 5e-5 | 0.840 | 0.824 | 0.824 | 0.818 |
| 2e-5 | 0.828 | 0.817 | 0.810 | 0.809 |
| Commit Message (CM) | | | | |
| 1e-3 | **0.880** | 0.867 | 0.876 | 0.760 |
| 5e-5 | 0.856 | 0.843 | 0.857 | 0.863 |
| 2e-5 | 0.829 | 0.844 | 0.862 | 0.797 |

The BERT-base-uncased model underwent pre-training using a vast collection of English text, employing a self-supervised approach. The classifier comprises a sequential combination of layers: a linear layer mapping the BERT output to the hidden layer, a ReLU activation function to introduce non-linearity, and another linear layer mapping the hidden layer to the output layer. In order to optimize the pre-trained BERT-base-uncased (768 hidden units, 12 layers, and 110M parameters) model to categorize SATD, we employed the CrossEntropyLoss as the designated loss function for the multi-class classification task. The model underwent training using the AdamW optimizer, which is a modified version of the Adam optimizer that incorporates weight decay (epsilon). Table 2 shows the results of all experiments utilizing BERT+AugGPT that have been carried out.

Table 2. The performance comparison results for BERT+AugGPT (categorization)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Epsilon** | | | |
| **Learning rate** | **1e-6** | **1e-7** | **1e-8** | **1e-9** |
| Code Comment (CC) | | | | |
| 3e-5 | 0.869 | 0.869 | 0.874 | 0.873 |
| 5e-5 | 0.875 | 0.877 | **0.882** | 0.874 |
| 2e-5 | 0.858 | 0.861 | 0.857 | 0.859 |
| Issue Section (IS) | | | | |
| 3e-5 | 0.860 | 0.860 | 0.840 | 0.835 |
| 5e-5 | 0.870 | 0.879 | **0.899** | 0.886 |
| 2e-5 | 0.820 | 0.822 | 0.844 | 0.866 |
| Pull Section (PS) | | | | |
| 3e-5 | 0.868 | 0.868 | 0.873 | 0.868 |
| 5e-5 | 0.863 | 0.868 | **0.876** | 0.873 |
| 2e-5 | 0.863 | 0.868 | 0.862 | 0.847 |
| Commit Message (CM) | | | | |
| 3e-5 | 0.840 | 0.847 | 0.837 | 0.830 |
| 5e-5 | 0.846 | 0.846 | **0.847** | 0.844 |
| 2e-5 | 0.834 | 0.838 | 0.838 | 0.837 |

[1] Srivastava. N.. Hinton. G.. Krizhevsky. A.. Sutskever. I.. & Salakhutdinov. R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*. *15*(1). 1929-1958.