

Sentiment Analysis with Deep Learning

Name: Ezgi Subaşı

Student ID: 20050111016

1. Preprocessing Decisions

Contraction Expansion: Replaced contractions (e.g., "don't" → "do not") for consistent text representation.

Email, URL, and HTML Tag Removal: Removed irrelevant elements using regex to retain only meaningful content.

Special Characters, Numbers, and Accented Characters: Eliminated non-alphabetic characters and converted accented ones to their non-accented forms for compatibility with GloVe embeddings.

Text Normalization: Converted text to lowercase, removed extra whitespaces, and corrected spelling errors for uniform tokenization.

2. Model Architecture

Embedding Layer: Used pre-trained 100-dimensional GloVe embeddings; words not in the GloVe vocabulary were assigned zero vectors.

Bidirectional LSTM Layers: Captured context in both directions for improved understanding of sentiment nuances.

Dropout Layers: Reduced overfitting by randomly deactivating neurons during training.

Dense Layers: Performed feature extraction and dimensionality reduction before the final classification.

Output Layer: Utilized softmax activation for three-class sentiment classification (negative, neutral, positive).

3. Analysis of Results

Training and Validation: Trained the model for 10 epochs with early stopping (patience=3) to prevent overfitting.

```
localhost:8888/notebooks/Desktop/NLPAssignment/SentimentAnalysisWithDeepLearning.ipynb
Jupyter SentimentAnalysisWithDeepLearning Last Checkpoint: 21 hours ago
File Edit View Run Kernel Settings Help
JupyterLab Python 3 (ipykernel)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in c:\users\sezgi\anaconda3\lib\site-packages (from rich->keras>=3.5.0->tensorflow-intel==2.18.0->tensorflow) (2.15.1)
Requirement already satisfied: mdurl==0.1 in c:\users\sezgi\anaconda3\lib\site-packages (from markdown-it-py>=2.2.0->rich->keras>=3.5.0->tensorflow-intel==2.18.0->tensorflow) (0.1.0)
Epoch 1/10
625/625 71s 103ms/step - accuracy: 0.5025 - loss: 2.1734 - val_accuracy: 0.6574 - val_loss: 0.9086
Epoch 2/10
625/625 58s 92ms/step - accuracy: 0.6888 - loss: 0.8036 - val_accuracy: 0.6814 - val_loss: 0.7788
Epoch 3/10
625/625 56s 90ms/step - accuracy: 0.7345 - loss: 0.6547 - val_accuracy: 0.6914 - val_loss: 0.7764
Epoch 4/10
625/625 57s 91ms/step - accuracy: 0.7863 - loss: 0.5594 - val_accuracy: 0.6756 - val_loss: 0.7673
Epoch 5/10
625/625 59s 94ms/step - accuracy: 0.8237 - loss: 0.4817 - val_accuracy: 0.6802 - val_loss: 0.8749
Epoch 6/10
625/625 59s 94ms/step - accuracy: 0.8588 - loss: 0.3958 - val_accuracy: 0.6822 - val_loss: 0.8895
Epoch 7/10
625/625 60s 95ms/step - accuracy: 0.8889 - loss: 0.3293 - val_accuracy: 0.6634 - val_loss: 0.9421
Epoch 8/10
196/196 5s 21ms/step
precision recall f1-score support
0 0.65 0.63 0.64 1824
1 0.57 0.62 0.59 2322
2 0.74 0.69 0.72 2101
accuracy 0.65 6247
macro avg 0.65 0.65 0.65 6247
weighted avg 0.65 0.65 0.65 6247
Confusion Matrix:
[[1145 571 108]
 [ 488 1444 390]
 [ 123 530 1448]]
```

Challenges and Improvements:

Initial simple architecture resulted in low accuracy (~41%).

Introducing bidirectional LSTMs and improved preprocessing significantly enhanced performance.

Used Keras Tuner to optimize hyperparameters within three hours. Find best parameters in below.

```
Trial 25 Complete [00h 06m 02s]
val_accuracy: 0.6998198628425598

Best val_accuracy So Far: 0.7050229907035828
Total elapsed time: 03h 03m 04s

Search: Running Trial #26

Value      |Best Value So Far|Hyperparameter
False      |True              |trainable_embeddings
256        |64                |lstm_units
0.2         |0.3               |dropout_rate
96          |96                |gru_units
64          |256               |dense_units
0.2         |0.4               |dense_dropout_rate
0.001      |0.0005           |learning_rate
10          |4                 |tuner/epochs
4           |0                 |tuner/initial_epoch
1           |1                 |tuner/bracket
1           |0                 |tuner/round
0019       |None              |tuner/trial_id

Epoch 5/10
```

Observations:

The model struggled with ambiguous or context-dependent texts. Future improvements could include attention mechanisms or fine-tuning transformer models like BERT.