Abstract:
The aim of this assignment is to provide an overview of the Word Embedding technique and its applications in Natural Language Processing (NLP). In this report, we will explore the concepts of Word2Vec, SVD Skip-Gram models, as well as their training techniques, and the most popular algorithms used to generate word embeddings. We will also demonstrate how to use pre-trained word embeddings in NLP tasks such as text classification and sentiment analysis. Finally, we will provide a comparison between different Word Embedding models, discussing their advantages and disadvantages.

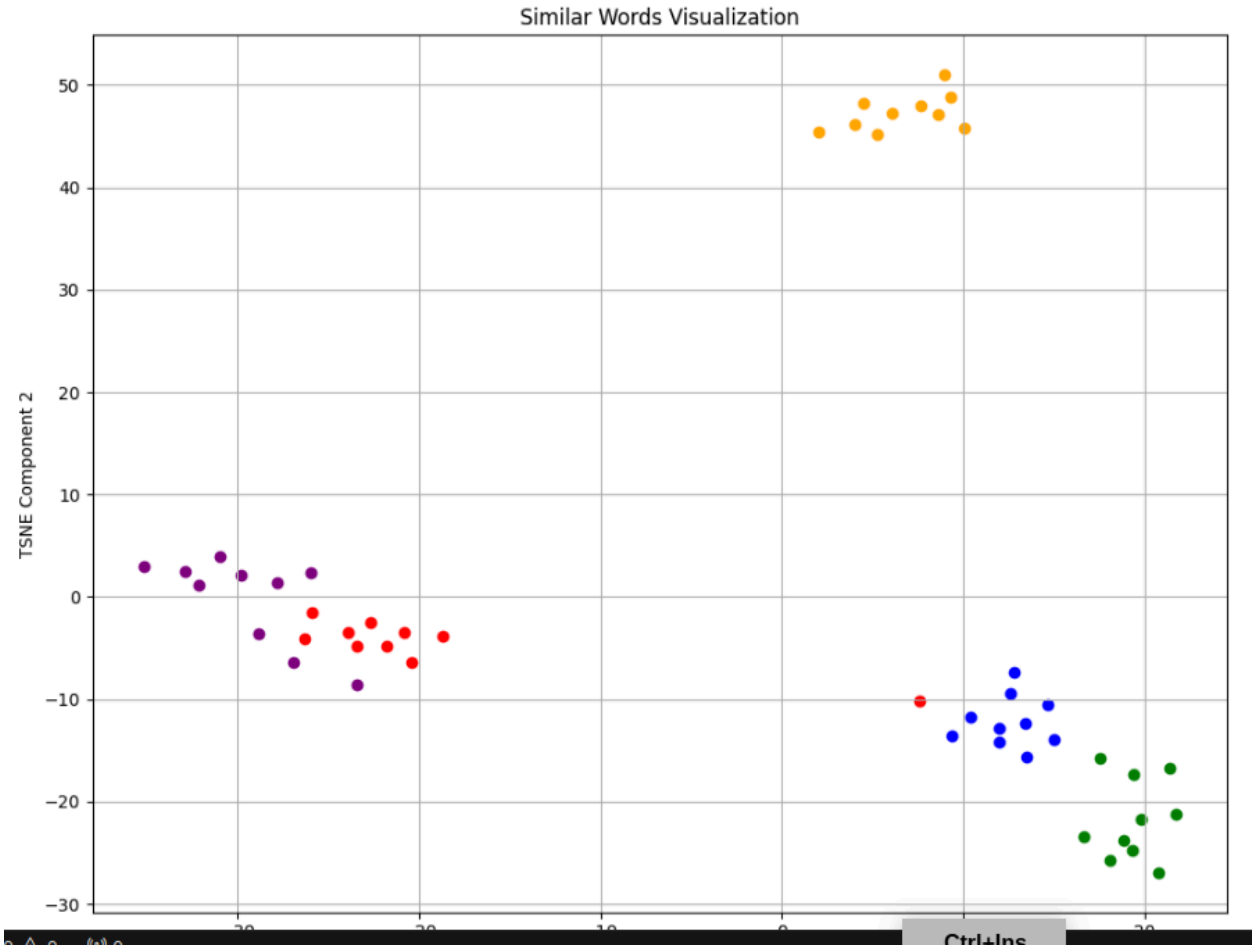
Results:
Two different approaches were taken to train word embeddings: the first method involved building a co-occurrence matrix and applying Singular Value Decomposition (SVD) to obtain word embeddings, while the second method used the CBOW model with Negative Sampling to train word vectors.

For the co-occurrence matrix approach, the

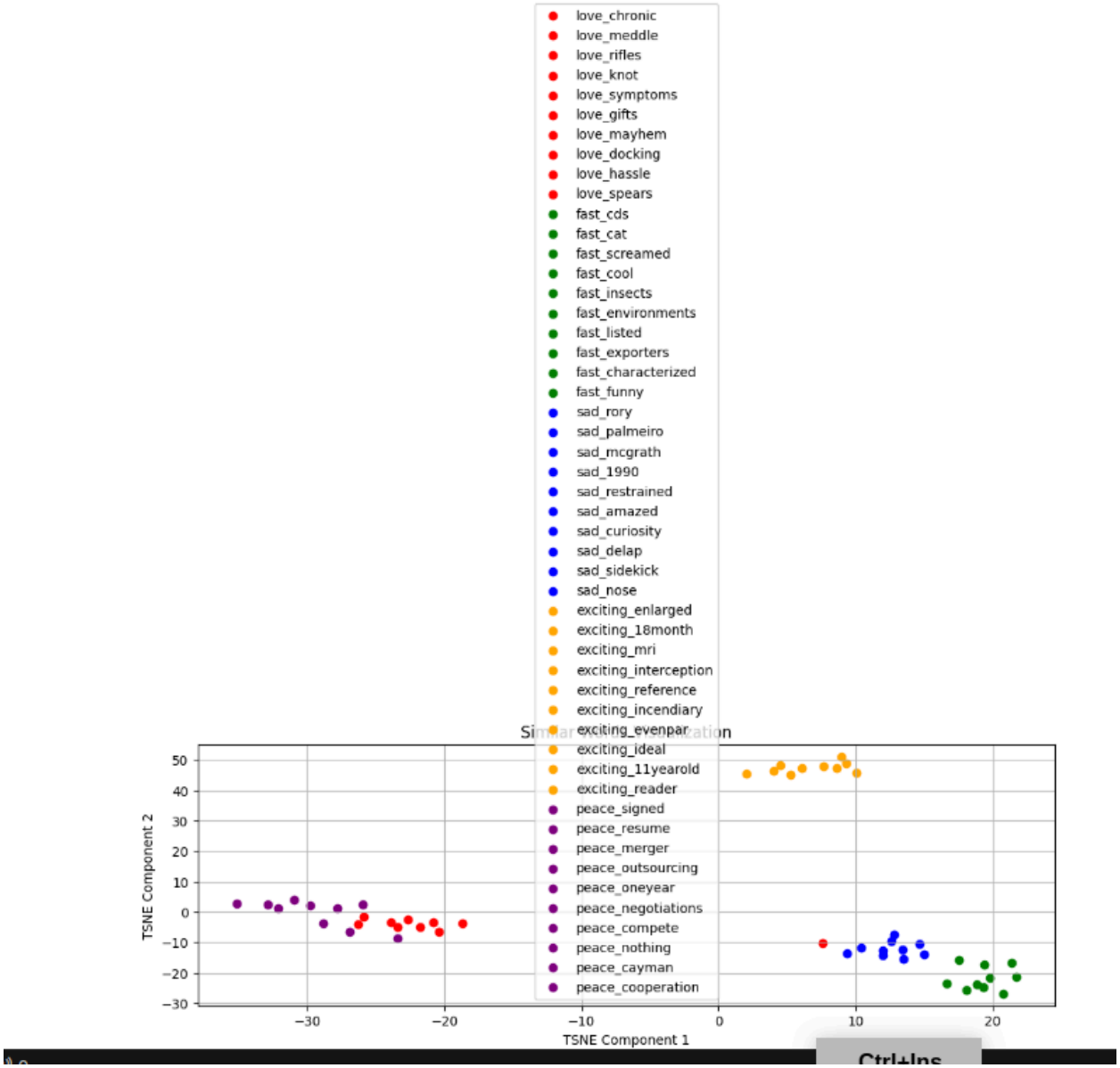
```
Top 10 similar words to the input word: natural
tanker
oao
volatile
exporter
skyhigh
exploration
threesession
lukoil
mix
trucks
```

Top-10 word vectors for five different words using t-SNE on a 2D plot for word2vec model

```
color_map = {
    'love': 'red',
    'fast': 'green',
    'sad': 'blue',
    'exciting': 'orange',
    'peace': 'purple'
}
```



Centre word - context word are here



SVD classification

Used Bi-LSTM for achieving more accuracy

```
num_sentences = 40001
```

```
learning_rate = 0.001
# Initialize the model and move it to GPU if available
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
model = BiLSTMClassifier(embedding_dim=100, hidden_dim=128, num_classes=5).to(device)
criterion = nn.CrossEntropyLoss()
optimizer = optim.Adam(model.parameters(), lr=learning_rate)

num_epochs = 5
```

```
    _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
  Epoch 1/5, Loss: 1.1073, Accuracy: 0.5010, Precision: 0.3975, Recall: 0.3989, F1 Score: 0.3964
  Epoch 2/5, Loss: 0.7359, Accuracy: 0.7085, Precision: 0.7040, Recall: 0.7067, F1 Score: 0.7045
  Epoch 3/5, Loss: 0.5979, Accuracy: 0.7807, Precision: 0.7793, Recall: 0.7798, F1 Score: 0.7788
  Epoch 4/5, Loss: 0.5448, Accuracy: 0.8034, Precision: 0.8025, Recall: 0.8028, F1 Score: 0.8021
  Epoch 5/5, Loss: 0.5056, Accuracy: 0.8172, Precision: 0.8164, Recall: 0.8166, F1 Score: 0.8160
  CPU times: user 13min 34s, sys: 9.01 s, total: 13min 43s
  Wall time: 13min 59s
```

Test accuracy = 67 %

```
Test Loss: 0.8501, Test Accuracy: 0.6704
Precision: 0.6816, Recall: 0.6704, F1 Score: 0.6650
Confusion Matrix:
[[1016  410  273  201]
 [  76 1638   85  101]
 [ 101  168 1368  263]
 [  98  239  489 1070]]
```