

EX.NO.: 17

DATE: 14.03.2025

FINDING SIMILAR VECTOR USING t-SNE

To visualize high-dimensional word embeddings from GloVe (Global Vectors for Word Representation) using t-SNE (t-distributed Stochastic Neighbor Embedding) for understanding the semantic relationship between words in a 2D space.

PROCEDURE:

1. Load GloVe Embeddings
2. Select Words for Visualization
3. Apply t-SNE for Dimensionality Reduction
4. Plot the Word Embeddings

CODE AND OUTPUT

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.manifold import TSNE
from gensim.models import KeyedVectors

def load_glove_embeddings(glove_file, num_words=5000):
    embeddings = {}
    with open(glove_file, 'r', encoding='utf-8') as file:
        for i, line in enumerate(file):
            if i >= num_words:
                break
            parts = line.split()
            word = parts[0]
            vector = np.array(parts[1:], dtype=np.float32)
            embeddings[word] = vector
    return embeddings

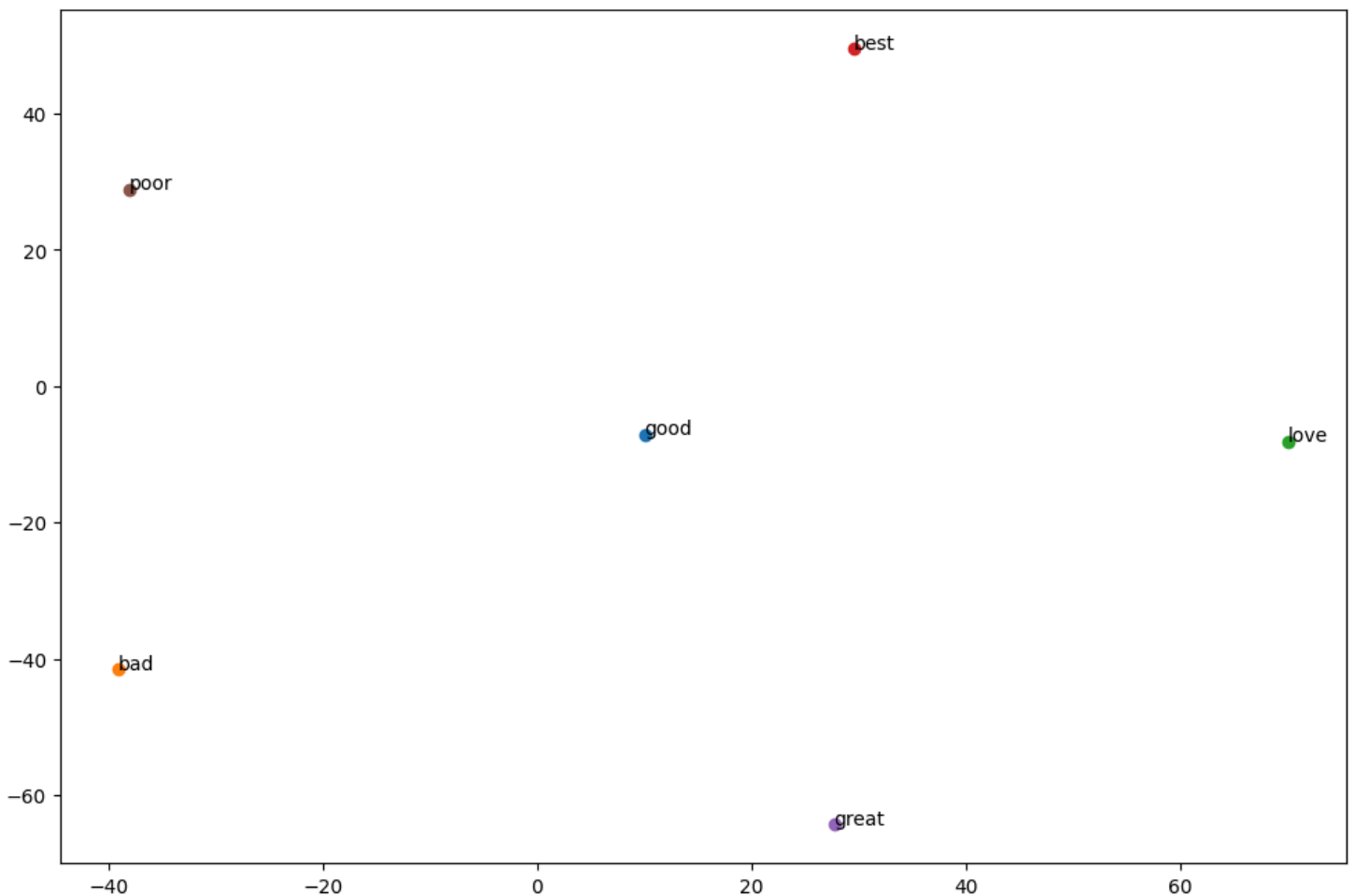
def visualize_embeddings(embeddings, words):
    vectors = np.array([embeddings[word] for word in words if word in embeddings])
    labels = [word for word in words if word in embeddings]

    tsne = TSNE(n_components=2, perplexity=min(30, len(vectors) - 1), random_state=42)
    reduced_vectors = tsne.fit_transform(vectors)

    plt.figure(figsize=(12, 8))
    for i, label in enumerate(labels):
        x, y = reduced_vectors[i]
        plt.scatter(x, y)
        plt.annotate(label, (x, y), fontsize=10)

    plt.show()
```

```
if __name__ == "__main__":  
    glove_path = "/content/glove.6B.50d.txt"  
    embeddings = load_glove_embeddings(glove_path, num_words=1000)  
  
    words_to_visualize = [  
        'good', 'bad', 'terrific', 'amazing', 'worst', 'dislike', 'fantastic',  
        'wonderful', 'nice', 'incredibly', 'love', 'hate', 'worse', 'best',  
        'okay', 'mediocre', 'excellent', 'awful', 'great', 'poor'  
    ]  
  
    visualize_embeddings(embeddings, words_to_visualize)
```



INFERENCE

Words with similar meanings and sentiments are grouped closer together, while words with opposite meanings are farther apart. This visualization helps in understanding the semantic relationships and emotional polarity between words in the GloVe vector space.