Name: Nilakshi Nagrale
Assignment: CSCI-544 Assignment 4 - Named Entity Recognition Using BiLSTM
USC ID: 2403347301

In this assignment, we had to focus on implementing Named Entity Recognition (NER) using deep learning models. We utilized the CoNLL-2003 corpus to train neural network models and performed evaluation on the development dataset. The assignment consists of three main tasks:

- implementing a simple bidirectional LSTM model (BLSTM)
- enhancing the model using GloVe embeddings
- optional bonus task that incorporates a CNN module.

Steps

Step 1: Understanding the Dataset

- The dataset contains three files: `train`, `dev`, and `test`.
- `train` and `dev` contain sentences with human-annotated NER tags, while `test` contains only raw sentences.
- Each line follows the format: `word_index word NER_tag` with sentences separated by blank lines.
- The **`glove.6B.100d.gz`** file provides pre-trained GloVe word embeddings.

Task 1: Simple Bidirectional LSTM Model

Implementation

- Implemented a BiLSTM network in PyTorch with the following architecture:
  Embedding → BLSTM → Linear → ELU → Classifier
- Hyper-parameters used:
  - Embedding Dimension: **100**
  - LSTM Layers: **1**
  - LSTM Hidden Dimension: **256**
  - Dropout: **0.33**
  - Linear Output Dimension: **128**
  - Optimizer: **SGD**
  - Learning Rate: **0.015**

- Batch Size: 16
- Epochs: 40

Evaluation on Development Data

- Precision: 81.64%
- Recall: 73.49%
- F1: 77.33%

processed 51578 tokens with 5942 phrases; found: 7029 phrases; correct: 5739.

accuracy: 92.35%; precision: 81.64%; recall: 73.49%; FB1: 77.33

LOC: precision: 81.84%; recall: 83.45%; FB1: 82.64 2134

MISC: precision: 75.47%; recall: 73.10%; FB1: 74.27 999

ORG: precision: 50.27%; recall: 69.43%; FB1: 58.32 2312

PER: precision: 87.59%; recall: 66.72%; FB1: 75.69 1584

Task 2: Using GloVe Word Embeddings

Enhancement Strategy

- Initialized the word embeddings in the model using GloVe vectors.
- Handled case-sensitivity issues since GloVe is case-insensitive.
- Maintained capitalization as it is crucial for NER.
- Retrained the BLSTM model with GloVe initialization.

Evaluation on Development Data

processed 51578 tokens with 5942 phrases; found: 7188 phrases; correct: 5918.

accuracy: 94.66%; precision: 92.36%; recall: 82.77%; FB1: 87.30

LOC: precision: 91.48%; recall: 90.04%; FB1: 90.15 2451

MISC: precision: 85.33%; recall: 75.27%; FB1: 79.77 1386

ORG: precision: 82.55%; recall: 76.06%; FB1: 79.55 1510

PER: precision: 95.21%; recall: 84.15%; FB1: 89.34 2341

Observations:

- GloVe embeddings capture word meanings and relationships, improving the model's contextual comprehension.
- Since GloVe is pre-trained on diverse datasets, it helps the model generalize effectively across different domains.
- With a large vocabulary, GloVe helps manage out-of-vocabulary words, leading to better overall performance.
- This results in better performance as compared to BiLSTM without GloVe embeddings.

Bonus: LSTM-CNN Model

Implementation

- Integrated a CNN module into the BLSTM model to capture character-level information.
- Hyper-parameters tuned:
    - Character Embedding Dimension: **30**
    - Number of CNN Layers: Tuned
    - Kernel Size: Tuned
    - Output Dimension per Layer: Tuned

Evaluation on Development Data

processed 51578 tokens with 5942 phrases; found: 6092 phrases; correct: 5892.
accuracy: 97.54%; precision: 96.73%; recall: 86.62%; FB1: 91.40
LOC: precision: 96.48%; recall: 93.74%; FB1: 95.09 2003
MISC: precision: 91.52%; recall: 81.13%; FB1: 85.98 885
ORG: precision: 90.12%; recall: 75.47%; FB1: 82.09 1260
PER: precision: 97.31%; recall: 90.39%; FB1: 93.73 1944

Conclusion

This assignment involved implementing and optimizing deep learning models for NER using the CoNLL-2003 dataset. We experimented with a simple BiLSTM model, improved it using GloVe word embeddings, and further enhanced performance with an LSTM-CNN architecture. The performance was evaluated using precision, recall, and F1 scores. The final models and predictions are saved as required, and required files are submitted as well.