Project Title:

Train and evaluate for named entity recognition using RNNs.

First Model Description:

The Sequential model is built with an Embedding layer, which is initialized with pre-trained word embeddings and converts integer-encoded vocabulary indices into dense vectors of fixed size. Following the Embedding layer is a SimpleRNN layer with 10 units, using the rectified linear unit (ReLU) activation function, and configured to return sequences. Subsequently, a TimeDistributed layer is added, applying the same Dense layer with 50 units to each time step of the sequence independently, using ReLU activation. The model concludes with an output Dense layer employing the softmax activation function, producing final outputs based on the specified number of tags (tag_len).

Second Model Description:

The Sequential model (m_2) is constructed with an Embedding layer, initialized with pre-trained word embeddings, converting integer-encoded vocabulary indices into dense vectors of a fixed size. Following the Embedding layer is a Bidirectional SimpleRNN layer with 10 units, configured to return sequences and using the rectified linear unit (ReLU) activation function. Subsequently, a TimeDistributed layer is added, applying the same Dense layer with 50 units to each time step of the sequence independently, using the ReLU activation function. The model concludes with an output Dense layer employing the softmax activation function, generating final outputs based on the specified number of tags (tag_len). The Bidirectional layer allows information to be processed in both forward and backward directions, enhancing the model's ability to capture context in the input sequences.

- ❖ A Bidirectional Recurrent Neural Network (Bidirectional RNN) is a type of neural network architecture that processes input sequences in both forward and backward directions. In the context of Natural Language Processing (NLP) and sequence modeling tasks, Bidirectional RNNs have been widely used to capture contextual information from both preceding and succeeding elements in a sequence.

  key points about Bidirectional RNNs:

  - Forward and Backward Processing:
    - ✓ Traditional RNNs, like SimpleRNN, process sequences from left to right (forward).
    - ✓ Bidirectional RNNs process sequences in both directions: from the beginning to the end (forward) and from the end to the beginning (backward).
  - Capturing Context:
    - ✓ Bidirectional processing helps capture contextual information from both past and future elements in the sequence.
    - ✓ This is particularly beneficial when understanding the meaning of a word or token depends on the context of both preceding and succeeding words.
  - Architecture:
    - ✓ In a Bidirectional RNN, two separate RNN layers (one processing forward and the other backward) are run in parallel.

- ✓ The outputs from both directions are typically concatenated or combined before being passed to the next layer.
- Use Cases:
  - ✓ Bidirectional RNNs are effective in tasks where understanding the entire sequence context is crucial, such as sentiment analysis, named entity recognition, and machine translation.
- Trade-offs:
  - ✓ While Bidirectional RNNs can capture more context, they are computationally more expensive and may introduce delays due to the need to process the entire sequence before generating an output.

Comparison:

SimpleRNN vs. Bidirectional SimpleRNN:

- The first model uses a standard SimpleRNN layer, processing sequences in a single direction.
- The second model enhances the RNN layer with Bidirectional, allowing it to capture information in both directions.
- <u>Consideration</u>: Bidirectional layers can be advantageous in capturing more comprehensive context, especially in tasks where context matters in both past and future directions.

Evaluation results:

SimpleRNN	Bidirectional SimpleRNN
All entities combined: Total entities: 18718 Total predicted: 17735 Correctly extracted: 13703 Precision: 77.27 % Recall: 73.21 % F-measure: 75.18 %	All entities combined: Total entities: 18718 Total predicted: 17660 Correctly extracted: 14262 Precision: 80.76 % Recall: 76.19 % F-measure: 78.41 %
Entity type: geo Total entities: 6194 Total predicted: 6453 Correctly extracted: 5123 Precision: 79.39 % Recall: 82.71 % F-measure: 81.02 % Entity type: gpe Total entities: 2757 Total predicted: 2753 Correctly extracted: 2536 Precision: 92.12 % Recall: 91.98 % F-measure: 92.05 %	Entity type: geo Total entities: 6194 Total predicted: 6532 Correctly extracted: 5292 Precision: 81.02 % Recall: 85.44 % F-measure: 83.17 % Entity type: gpe Total entities: 2757 Total predicted: 2676 Correctly extracted: 2552 Precision: 95.37 % Recall: 92.56 % F-measure: 93.94 %

SimpleRNN	Bidirectional SimpleRNN
Entity type: per Total entities: 2784 Total predicted: 2626 Correctly extracted: 1848 Precision: 70.37 % Recall: 66.38 % F-measure: 68.32 % Entity type: org Total entities: 3400 Total predicted: 2807 Correctly extracted: 1685 Precision: 60.03 % Recall: 49.56 % F-measure: 54.29 % Entity type: tim Total entities: 3431 Total predicted: 3096 Correctly extracted: 2511 Precision: 81.1 % Recall: 73.19 % F-measure: 76.94 % Entity type: art Total entities: 75 Total predicted: 0 Correctly extracted: 0 Precision cannot be computed. Recall: 0.0 % F-measure cannot be computed. Entity type: nat Total entities: 36 Total predicted: 0 Correctly extracted: 0 Precision cannot be computed. Recall: 0.0 % F-measure cannot be computed. Entity type: eve Total entities: 41 Total predicted: 0 Correctly extracted: 0 Precision cannot be computed. Recall: 0.0 % F-measure cannot be computed.	Entity type: per Total entities: 2784 Total predicted: 2525 Correctly extracted: 1887 Precision: 74.73 % Recall: 67.78 % F-measure: 71.09 % Entity type: org Total entities: 3400 Total predicted: 2863 Correctly extracted: 1952 Precision: 68.18 % Recall: 57.41 % F-measure: 62.33 % Entity type: tim Total entities: 3431 Total predicted: 3064 Correctly extracted: 2579 Precision: 84.17 % Recall: 75.17 % F-measure: 79.41 % Entity type: art Total entities: 75 Total predicted: 0 Correctly extracted: 0 Precision cannot be computed. Recall: 0.0 % F-measure cannot be computed. Entity type: nat Total entities: 36 Total predicted: 0 Correctly extracted: 0 Precision cannot be computed. Recall: 0.0 % F-measure cannot be computed. Entity type: eve Total entities: 41 Total predicted: 0 Correctly extracted: 0 Precision cannot be computed. Recall: 0.0 % F-measure cannot be computed.

Comments on Evaluation:

SimpleRNN
Training Metrics (Epochs 1/2 and 2/2): • Loss: The loss is decreasing over the epochs, which is a positive sign. It means that your model is improving in minimizing the difference between predicted and actual values. • Accuracy: The accuracy is increasing, reaching 97.23% in the second epoch. This suggests that your model is performing well in terms of correctly classifying instances.
Bidirectional SimpleRNN

Training Metrics (Epochs 1/2 and 2/2): • Loss: The loss is decreasing, indicating that the model is effectively learning from the training data. • Accuracy: The accuracy is improving, reaching 97.67% in the second epoch, suggesting strong predictive performance.
Comparison
Precision: • The Bidirectional SimpleRNN model outperforms the SimpleRNN model in precision, with 81.39% compared to 79.01%. This suggests that the Bidirectional model is better at avoiding false positives. Recall: • The Bidirectional SimpleRNN model also performs better in recall, with 85.34 %compared to 79.01 %.This indicates that the Bidirectional model is better at capturing more of the actual entities present in the data. F-measure: • The F-measure, which balances precision and recall, is higher for the Bidirectional SimpleRNN model (83.32 %)compared to the SimpleRNN model (80.52 %)This indicates an overall better performance in terms of both false positives and false negatives. Conclusion: • Based on these results, it seems that the Bidirectional SimpleRNN model is more effective in capturing the underlying patterns in the data for the named entity recognition task compared to the SimpleRNN model. The bidirectional nature of the model allows it to consider information from both past and future time steps, enhancing its ability to recognize entities within the sequences.

Error Analysis:

SimpleRNN	
While the model performs well for some entity types (geo, gpe), there are areas for improvement, especially in recognizing persons (per), organizational entities (org), and time-related entities (tim). Additionally, the model did not predict any entities for art, nat, and eve, indicating potential challenges in these specific categories. Fine-tuning the model or considering additional training data may enhance performance, especially for the underperforming entity types.	
Bidirectional SimpleRNN	
While the model performs reasonably well for time-related entities, there are areas for improvement, especially in recognizing organizational entities. The model struggles to predict entities for art, nat, and eve categories, suggesting potential challenges in these specific entity types that require further investigation and refinement. Adjustments to the model or the inclusion of additional data might help improve performance in these challenging areas.	
Comparison	
SimpleRNN	**Bidirectional SimpleRNN**
Sentence: ['General', 'Halutz', 'submitted', 'his', 'resignation', 'to', 'Israeli', 'Prime', 'Minister', 'Ehud', 'Olmert', 'and', 'Defense', 'Minister', 'Amir', 'Peretz', 'Tuesday', '.'] Target: ['B-per', 'I-per', 'O', 'O', 'O', 'O', 'B-gpe', 'B-per', 'O', 'B-per', 'I-per', 'O', 'O', 'O', 'B-per', 'I-per', 'B-tim', 'O'] Predicted: ['B-per', 'I-per', 'O', 'O', 'O', 'O', 'B-gpe', 'B-per', 'I-per', 'B-per', 'I-per', 'O', 'O', 'O', 'I-org', 'I-per', 'B-tim', 'O']	Sentence: ['General', 'Halutz', 'submitted', 'his', 'resignation', 'to', 'Israeli', 'Prime', 'Minister', 'Ehud', 'Olmert', 'and', 'Defense', 'Minister', 'Amir', 'Peretz', 'Tuesday', '.'] Target: ['B-per', 'I-per', 'O', 'O', 'O', 'O', 'B-gpe', 'B-per', 'O', 'B-per', 'I-per', 'O', 'O', 'O', 'B-per', 'I-per', 'B-tim', 'O'] Predicted: ['B-per', 'I-org', 'O', 'O', 'O', 'O', 'B-gpe', 'B-per', 'I-per', 'I-per', 'I-per', 'O', 'I-org', 'O', 'I-per', 'I-per', 'B-tim', 'O']

At 0 ('per', 'General Halutz') Extracted. At 6 ('gpe', 'Israeli') Extracted. At 7 ('per', 'Prime') Missed. At 9 ('per', 'Ehud Olmert') Extracted. At 14 ('per', 'Amir Peretz') Missed. At 16 ('tim', 'Tuesday') Extracted. At 7 ('per', 'Prime Minister') Incorrectly extracted.	At 0 ('per', 'General Halutz') Missed. At 6 ('gpe', 'Israeli') Extracted. At 7 ('per', 'Prime') Missed. At 9 ('per', 'Ehud Olmert') Missed. At 14 ('per', 'Amir Peretz') Missed. At 16 ('tim', 'Tuesday') Extracted. At 0 ('per', 'General') Incorrectly extracted. At 7 ('per', 'Prime Minister Ehud Olmert') Incorrectly extracted.
Sentence: ['Hurricane', 'Dean', 'struck', 'the', 'island', 'in', 'August', '2007', 'causing', 'damages', 'equivalent', 'to', '20', '%', 'of', 'GDP', '.'] Target: ['B-eve', 'I-eve', 'O', 'O', 'O', 'O', 'B-tim', 'I-tim', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O'] Predicted: ['O', 'B-per', 'O', 'O', 'O', 'O', 'B-tim', 'I-tim', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O'] At 0 ('eve', 'Hurricane Dean') Missed. At 6 ('tim', 'August 2007') Extracted. At 1 ('per', 'Dean') Incorrectly extracted.	Sentence: ['Hurricane', 'Dean', 'struck', 'the', 'island', 'in', 'August', '2007', 'causing', 'damages', 'equivalent', 'to', '20', '%', 'of', 'GDP', '.'] Target: ['B-eve', 'I-eve', 'O', 'O', 'O', 'O', 'B-tim', 'I-tim', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O'] Predicted: ['O', 'B-per', 'O', 'O', 'O', 'O', 'B-tim', 'I-tim', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O'] At 0 ('eve', 'Hurricane Dean') Missed. At 6 ('tim', 'August 2007') Extracted. At 1 ('per', 'Dean') Incorrectly extracted.
Sentence: ['President', 'Bush', 'says', 'federal', 'officials', 'have', '"', 'deep', 'concern', '"', 'about', 'Tropical', 'Storm', 'Rita', 'causing', 'more', 'flooding', 'in', 'New', 'Orleans', '.'] Target: ['B-per', 'I-per', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-nat', 'I-nat', 'I-nat', 'O', 'O', 'O', 'O', 'B-geo', 'I-geo', 'O'] Predicted: ['B-per', 'I-per', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'I-org', '_PAD_', 'O', 'O', 'O', 'O', 'B-geo', 'I-geo', 'O'] At 0 ('per', 'President Bush') Extracted. At 11 ('nat', 'Tropical Storm Rita') Missed. At 18 ('geo', 'New Orleans') Extracted.	Sentence: ['President', 'Bush', 'says', 'federal', 'officials', 'have', '"', 'deep', 'concern', '"', 'about', 'Tropical', 'Storm', 'Rita', 'causing', 'more', 'flooding', 'in', 'New', 'Orleans', '.'] Target: ['B-per', 'I-per', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-nat', 'I-nat', 'I-nat', 'O', 'O', 'O', 'O', 'B-geo', 'I-geo', 'O'] Predicted: ['B-per', 'I-per', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'I-org', 'I-org', 'O', 'O', 'O', 'O', 'B-geo', 'I-geo', 'O'] At 0 ('per', 'President Bush') Extracted. At 11 ('nat', 'Tropical Storm Rita') Missed. At 18 ('geo', 'New Orleans') Extracted.
Sentence: ['Iranian', 'Foreign', 'Ministry', 'spokesman', 'Hamid', 'Reza', 'Asefi', 'made', 'that', 'assertion', 'Sunday', 'at', 'his', 'weekly', 'news', 'conference', 'in', 'Tehran', '.'] Target: ['B-gpe', 'B-org', 'I-org', 'O', 'B-per', 'I-per', 'I-per', 'O', 'O', 'O', 'B-tim', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'O'] Predicted: ['B-gpe', 'O', 'I-org', 'O', 'B-per', 'I-per', 'I-per', 'O', 'O', 'O', 'B-tim', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'O'] At 0 ('gpe', 'Iranian') Extracted. At 1 ('org', 'Foreign Ministry') Missed. At 4 ('per', 'Hamid Reza Asefi') Extracted. At 10 ('tim', 'Sunday') Extracted. At 17 ('geo', 'Tehran') Extracted.	Sentence: ['Iranian', 'Foreign', 'Ministry', 'spokesman', 'Hamid', 'Reza', 'Asefi', 'made', 'that', 'assertion', 'Sunday', 'at', 'his', 'weekly', 'news', 'conference', 'in', 'Tehran', '.'] Target: ['B-gpe', 'B-org', 'I-org', 'O', 'B-per', 'I-per', 'I-per', 'O', 'O', 'O', 'B-tim', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'O'] Predicted: ['B-gpe', 'B-org', 'I-org', 'O', 'B-per', 'I-per', 'I-per', 'O', 'O', 'O', 'B-tim', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'O'] At 0 ('gpe', 'Iranian') Extracted. At 1 ('org', 'Foreign Ministry') Extracted. At 4 ('per', 'Hamid Reza Asefi') Extracted. At 10 ('tim', 'Sunday') Extracted. At 17 ('geo', 'Tehran') Extracted.
Sentence: ['Roche', 'also', 'announced', 'that', 'it', 'would', 'provide', 'Taiwan', 'with', 'an', 'additional', '1.3', 'million', 'treatments', 'of', 'Tamiflu', '.'] Target: ['B-org', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'O'] Predicted: ['B-per', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'O'] At 0 ('org', 'Roche') Missed. At 7 ('geo', 'Taiwan') Extracted. At 15 ('geo', 'Tamiflu') Extracted. At 0 ('per', 'Roche') Incorrectly extracted.	Sentence: ['Roche', 'also', 'announced', 'that', 'it', 'would', 'provide', 'Taiwan', 'with', 'an', 'additional', '1.3', 'million', 'treatments', 'of', 'Tamiflu', '.'] Target: ['B-org', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'O'] Predicted: ['B-per', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'O'] At 0 ('org', 'Roche') Missed. At 7 ('geo', 'Taiwan') Extracted. At 15 ('geo', 'Tamiflu') Extracted. At 0 ('per', 'Roche') Incorrectly extracted.
• These two models struggles to predict entities for art, nat, and eve categories, suggesting potential challenges in these specific entity types that require further investigation and refinement.	

- Both models struggle with identifying multi-word entities, such as "Prime Minister Ehud Olmert" and "Tropical Storm Rita."
- Model 1 tends to incorrectly label entities, such as identifying "Israeli" as an organization.
- Model 2 has a tendency to miss entities, such as missing "General Halutz" and "Roche."
- Both models struggle with ambiguous entities, like "Dean," which could refer to a person or an event.

Result on my own sentence(using second model as the best model):

Evaluation results: Entity type: geo Total entities: 4 Total predicted: 3 Correctly extracted: 1 Precision: 33.33 % Recall: 25.0 % F-measure: 28.57 % Entity type: gpe Total entities: 0 Total predicted: 0 Correctly extracted: 0 Precision cannot be computed. Recall cannot be computed. F-measure cannot be computed. Entity type: per Total entities: 8 Total predicted: 2 Correctly extracted: 2 Precision: 100.0 % Recall: 25.0 % F-measure: 40.0 % Entity type: org Total entities: 1 Total predicted: 0 Correctly extracted: 0 Precision cannot be computed. Recall: 0.0 % F-measure cannot be computed. Entity type: tim Total entities: 1 Total predicted: 0 Correctly extracted: 0 Precision cannot be computed. Recall: 0.0 % F-measure cannot be computed. Entity type: art Total entities: 0 Total predicted: 0 Correctly extracted: 0 Precision cannot be computed. Recall cannot be computed. F-measure cannot be computed. Entity type: nat Total entities: 0 Total predicted: 0	**Entity Type: geo** Precision (33.33%): The model correctly predicted 33.33% of the entities for the 'geo' type. This suggests that there were some false positives in the predictions. Recall (25.0%): The model captured 25.0% of the actual 'geo' entities, indicating that there were false negatives, and some entities were not detected. **Entity Type: gpe** Precision, Recall, and F-measure (Not computable): Since there were no entities of type 'gpe' in the ground truth, these metrics cannot be computed. The model correctly refrained from making predictions for this type. **Entity Type: per** Precision (100.0%): The model achieved perfect precision for 'per,' indicating that all predicted entities were correct. Recall (25.0%): However, the recall is low, suggesting that many actual 'per' entities were missed by the model. **Entity Types: org, tim, art, nat, eve** Precision, Recall, and F-measure (Not computable): Similar to 'gpe,' these metrics cannot be computed as there were no entities of these types in the ground truth. **All Entities Combined** Precision (60.0%): The overall precision indicates that 60.0% of the predicted entities across all types were correct. Recall (21.43%): The recall is relatively low, indicating that a significant portion of actual entities was not captured by the model. F-measure (31.58%): The F-measure provides a balance between precision and recall, and the value suggests that there is room for improvement. **Comments:** The model seems to struggle with recall across multiple entity types, indicating that it misses some entities during prediction. High precision for 'per' suggests that the model is accurate when it predicts entities of this type, but the low recall means it misses many 'per' entities. The absence of predictions and perfect precision for types with no entities in the ground truth is a positive aspect, indicating that the model refrains from making false predictions in such cases.

Correctly extracted: 0 Precision cannot be computed. Recall cannot be computed. F-measure cannot be computed. Entity type: eve Total entities: 0 Total predicted: 0 Correctly extracted: 0 Precision cannot be computed. Recall cannot be computed. F-measure cannot be computed. All entities combined: Total entities: 14 Total predicted: 5 Correctly extracted: 3 Precision: 60.0 % Recall: 21.43 % F-measure: 31.58 %	